Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

A posteriori error estimation by means of the exactly equilibrated fields

I. Anufriev, V. Korneev, V. Kostylev

RICAM-Report 2007-08

A posteriori error estimation by means of the exactly equilibrated fields

I. Anufriev[2] , V. Korneev[1,2] and[*†] V. Kostylev[3]

November 17, 2007

Dedicated to L.A. Rozin

[1] St. Petersburg State University, Russia,
[2] St. Petersburg State Polytechnical University, Russia
[2] All-Russian Scientific-Research Institute of Hydroengineering, Russia

Anufriev@amd.stu.neva.ru, Korneev@amd.stu.neva.ru, Vol_Kos@mail.ru

Abstract

In this paper, we advocate the "classical" approach to the a posteriori error estimation, which for the theory elasticity problems stems from the Lagrange and Castigliano variational principles. In it, the energy of the error of an approximate solution, satisfying geometrical restrictions, is estimated by the energy of the difference of the stress tensor corresponding to the approximate solution and any stress tensor, satisfying the equations of equilibrium. Notwithstanding a popular point of view that the construction of equilibrated stress fields requires considerable computational effort, we show that it can be practically always done for a number of arithmetic operations, which is asymptotically optimal. Numerical experiments show that a posteriori error estimators, based on the use of exactly equilibrated stress fields, provide very good coefficients of effectiveness, which in many cases can be convergent to the unity. At the same time they have linear complexity and are robust.

1 Introduction

Publications on a posteriori error estimates for approximate, *e.g.,* numerical solutions of partial differential equations are numerous. The earliest a posteriori error estimates were apparently known in mechanics from the time when the Lagrange and Castigliano principles, which, from the mathematical point of view, provide the primal and dual formulations in theory elasticity,

[*] Research is supported by the grant from the Russian Fund of Basic Research N 05-01-00779-a.
[†] The second author has partially been supported by Johann Radon Institute for Computational and Applied Mathematics (RICAM) of the Austrian Academy of Sciences (AAS), Linz, Austria.

obtained a mature form. Such estimates are deduced from the fact that approximate solutions obtained on the basis of these principles approach the exact solution in the energy sense from the opposite directions, and, namely, from above and from below, respectively. Let $\mathbf{u}$ be the vector of exact displacements of a linearly elastic body, $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\mathbf{u})$ is the corresponding stress tensor, $\|\cdot\|_U$ and $|\cdot|_\sigma$ are the potential energy norms expressed in displacements and stresses. If $\widetilde{\mathbf{u}}$ is an arbitrary displacement vector of finite energy, satisfying geometric boundary conditions, and $\boldsymbol{\tau}$ is an arbitrary stress tensor of finite energy, satisfying the equilibrium equations (including the boundary conditions in stresses), then the classical a posteriori estimate is

$$\|\mathbf{u} - \widetilde{\mathbf{u}}\|_U \le |\boldsymbol{\sigma}(\widetilde{\mathbf{u}}) - \boldsymbol{\tau}|_\sigma \,, \tag{1.1}$$

and may be found, *e.g.*, in the Mikhlin's book [33]. In spite of its simple form and enormous amount of publications on a posteriori estimates, the authors were unable to find references where it had been directly used in practical FE applications.

During several last decades, a few groups of a posteriori error estimation techniques have been developed and, first of all, so called residual-based techniques found in Babuška I and Reinbolt [5, 6], Verfürth [51] Stewart and Hudges [48] and more recent publications. Among them there are distinguished the explicit residual method, see Bernardi and Girault [9] and Carstensen [11] and the related paper by Clement [12] on special interpolation, and the implicit and equilibrated residual methods, for which we refer to Babuška I and Reinbolt [5], Kelly [18], Ladevese and Leguillon [31], and Ainsworth, Demkowicz and Kim [1]. Many papers are dedicated to obtaining of indicators of the error, which do not necessarily bound it, but are approximately proportional to the error, and, therefore, can be used for the mesh refinement in adaptive computations. A pioneering paper of Zienkiewicz and Zhu [55], see also Ainsworth and Oden [2], commenced the group of such techniques widely used in applications and employing superconvergence properties of finite element solutions. Many contributions have been related to a posteriori error estimation for specific problems of mathematical physics. The development of a posteriori estimation techniques in recent decades, as well as the bibliography, are reflected in the books of Aubin [3], Verfürth [52], Ainsworth and Oden [2] Babuska and Strouboulis [7] and Neittaanmäki and Repin [35].

The main idea of some of the mentioned approaches to the error estimation is to use fields of stresses, which can be obtained from the FE solution and at the same time are most close to the exact equilibrated fields (*i.e.,* representing the exact solution of the problem). An example is the equilibrated residual method, which now gains more attention as a method which allow to obtain reliable bounds, often without solving some global systems of algebraic equations, see, *e.g.,* Ainsworth, Demkowicz and Kim [1], Luce and Wohlmuth [32], Vejchodský [50] and Braess and Schöberl [10]. However, it is also true that the purpose of most authors is to outflank construction of *exactly equilibrated fields* at all[1]. One way of obtaining equilibrated fields[2], which approximate equilibrated fields of the exact solution of the primal problem, is approximate solution of the dual problem, which in the theory of elasticity is expressed by the Castigliano principle of virtual equilibrated states. As a rule a motivation for avoiding the use of equilibrated fields is that

[1]In general discussions we use the theory elasticity problem for a model without special remarks.

[2]In the paper the both terms *equilibrated* and *exactly equilibrated field* imply that a field satisfies equilibrium equations exactly in classical or generalized sense.

the solution of the dual problem or other ways of finding such fields are computationally too expensive.

The purpose of this paper is to illustrate that in many cases the estimate (1.1) can be directly used as an *efficient and cheap* error estimator. This is for the reason that indeed equilibrated fields are not difficult to find in a variety of ways. As one of the options, as we will see not the most efficient in many cases, the numerical solution of the dual problem can be considered. For advocating this option, the following fact is important: numerical solutions of the discretizations of the primal and dual problems, having the same (in the order) accuracy in the energy norm, can be found for the same (in the order) computational cost. More over, under some conditions, the discretization of the dual problem may be obtained in such a way that its matrix will coincide with the FE matrix for the primal problem up to the boundary conditions. Therefore, practically the same solver can be used for solution of the both discrete problems.

From the above discussion, one concludes that the option of solving the dual problem for evaluating the equilibrated fields deserves examination. Suppose that discretizations of the same order of accuracy are used for the primal and dual problems. In general, one can expect that the efficiency coefficient will converge to the unity at $h \to 0$, if the error estimator is super-convergent. In the practice, such convergence was observed for a number of alternative a posteriori error estimators, considered in this paper and papers of other authors, see, *e.g.,* Luce and Wohlmuth [32]. In our numerical experiments with the use of classical error estimate of the type (1.1) and of the equilibrated fields, obtained by solution of the dual problem with the same accuracy, the observed effectiveness coefficient remained close to 1.25. Further improvement of the efficiency coefficient is possible, *e.g.,* if we solve the dual problem on a denser, then for the primal problem, mesh, which results in a greater cost of an error estimator, than the cost of the solution to be validated.

At least not less promising approach can be referred as *direct evaluation of the equilibrated fields*. It is based on the fact that the equilibrium equations (in stresses) are under-determined. For instance, in the theory of elasticity, the symmetric stress tensor $\boldsymbol{\sigma} = \{\sigma_{kl}\}_{k,l=1}^{3}$ with six stresses for the entries satisfies three equilibrium equations. Therefore, in order to satisfy the equilibrium equations it is sufficient to perform two steps:

1) *to specify three stresses, say shear stresses*

$$\sigma_{kl}\,, \qquad k \neq l, \tag{1.2}$$

by arbitrary sufficiently smooth functions and

2) to find the rest stresses from the equilibrium equations by evaluating 1-d integrals.

The presence of the boundary conditions for stresses often does not make this procedure significantly more difficult. When this procedure is used in the a posteriori estimator, *e.g.,* for the FE solution, the stresses (1.2) are found from the FE solution with the accurate use of its superconvergence properties.

The approach of direct evaluation of the equilibrated fields allows us not only to design cheap algorithms for evaluating a posteriori error estimators, often producing convergent the efficiency coefficients, but to come to a new type of general a posteriori estimates. For simplicity, let us consider the Dirichlet problem for the Poisson equation in the rectangle $\Omega = (a_1, b_1) \times (a_2, b_2)$

$$-\Delta u = f(x)\,, \qquad x = (x_1, x_2) \in \Omega\,, \qquad u\big|_{\partial\Omega} = 0\,, \tag{1.3}$$

and any function v from $\mathring{H}^1(\Omega) = \{w \in H^1(\Omega) : u\big|_{\partial\Omega} = 0\}$, which is considered as an "approximation" for u. Then for any $\epsilon > 0$ the error estimate can be written in the form

$$\begin{aligned} &\|\nabla(v-u)\|^2_{0,\Omega} \le \\ &\le (1+\epsilon)\|\nabla v - \mathbf{y}\|^2_{0,\Omega} + (1+\tfrac{1}{\epsilon})\|\textstyle\sum_{k=1,2}\int_{a_k(x_{3-k})}^{x_k} \alpha_k(f - \nabla\cdot\mathbf{y})(\eta_k, x_{3-k})\,d\eta_k\|^2_{0,\Omega}\,, \end{aligned} \tag{1.4}$$

$\mathbf{y} = (y_1, y_2)^\top$ is any sufficiently smooth vector-function, functions $\alpha_k = \alpha_k(x)$ satisfy $\alpha_1 + \alpha_2 \equiv 1$, and by convention $\nabla\cdot\mathbf{y} = \operatorname{div}\mathbf{y}$. If $f \in L_2(\Omega)$ and $\alpha_k \in C(\Omega)$, it is natural to consider $\mathbf{y} \in H(\Omega, \operatorname{div}) = \{\mathbf{z} : \nabla\cdot\mathbf{z} \in L_2(\Omega)\}$. If $v = u_{\rm fem}$ is the FE solution, then $\mathbf{y}$ may be obtained by the averaging of the derivatives $\partial u_{\rm fem}/\partial x_k$ at the nodes and interpolation. In our numerical experiments, algorithms of that type almost always produced good and convergent effectiveness coefficients. In the algorithms presented in the paper we implement a variety of techniques of direct evaluation of the exactly equilibrated fields for their use in the a posteriori estimators.

Clearly, estimate (1.4) fetches additional opportunities in comparison with the known estimates of a close appearance, in which the second term in the right part is, *e.g.*, $\|f - \nabla\cdot\mathbf{y}\|_{-1,\Omega}$. Numerical evaluation of this negative norm is not at all easy, whereas its replacement in (1.4) can be often computed for a number of arithmetic operations proportional to the number of unknowns. In particular, this is true for the FE discretizations by means of the orthogonal grids. Another example is Galerkin methods with the coordinate functions specified by analytical expressions in the whole domain. In the paper we discuss also the procedures, which can provide the optimal computational cost in more general situations.

The equilibrium equations in the theories of thin shells and shells of moderate thickness, see, *e.g.,* Gol'denweizer [17] Novozhilov [37, 36] and Reissner [44], are written in terms of internal forces, *i.g.,* shear forces and bending and twisting moments. These equations are more complicated, than the equilibrium equations in the theory of elasticity. However, a quite similar approach can be implemented for obtaining the exactly equilibrated functions of internal forces. This approach originates from the papers devoted to numerical methods for solving the bending thin plate and shell problems on the basis of the Castigliano principle and the method of splitting of the thin plate and shell partial differential equations, studied by Rozin [45, 46], Korneev/Rozin [29, 30], Korneev [19, 20, 21, 25, 27].

Considerable part of the algorithms contained in the paper were tested numerically. We present the graphs and tables of numbers illustrating the dependence of effectiveness indices and arithmetical work on numbers of unknowns. Obviously, all tested algorithms are optimal in the computational work. Additionally to this, the practical computational costs of the a posteriori error estimators and of the optimal multigrid solvers for the primal problem were compared. As a rule the latter exceeded the former in two times at the least.

The paper is arranged as follows. In Section 2, we consider a posteriori estimators for the Poisson equation in the unite square and arbitrary Lipshits continuous domain with different boundary conditions. The special case of the differential operator with the discontinuous coefficient in the main term is treated in Subsection 2.3. It is shown that the algorithms of a posteriori estimators are easily adjusted to this case. The results of numerical tests for this case, presented in Subsection 5.2, show that discontinuity practically does not affect the efficiency of the a posteriori error estimator. Section 3 deals with the a posteriori error estimators for the

plane elasticity problem. All a posteriori error estimators of Sections 2 and 3 are based on the *direct evaluation* of the balanced fluxes and equilibrated stress tensors, *i.e.,* without solving any systems of algebraic equations. In Section 4.1, we consider an alternative approach based on solution of the dual problem equivalent to the Castigliano principle of virtual complementary work. We show that it is possible to chose a basis in the space of the self-equilibrated tensors of stresses in such a way that the system of algebraic equations will possess practically the same properties as the FE system for the primal problem. Throughout the paper alongside with the general algorithms, we present the algorithms, which we tested numerically. Results of numerical experiments are discussed in Section 5.

Throughout the paper, we use the notations listed below.

$\mathcal{P}_p$ and $\mathcal{Q}_p$ are the spaces of polynomials of the total order p and of the order p in each variable, $\mathbf{e}_1 = (1,0)$, $\mathbf{e}_2 = (0,1)$, d is the dimension.

$\mathbf{L}_2(\Omega)$ is the space $[L_2(\Omega)]^d$ with the norm $\|\cdot\| = \|\cdot\|_{0,\Omega}$ and the same notation is used for the norm in $L_2(\Omega)$,

$|\cdot|_{k,\Omega}$, $\|\cdot\|_{k,\Omega}$ stand for the semi-norm and the norm in the Sobolev space $H^k(\Omega)$, *i.e.*,

$$| v |^2_{k,\Omega} = \sum_{|q|=k} \int_\Omega (D^q_x v)^2 d\mathbf{x}\,, \qquad \| v \|^2_{k,\Omega} = \|v\|^2_{0,\Omega} + \sum_{l=1}^{k} |v|^2_{l,\Omega}\,,$$

where

$$D^q_x v := \partial^{|q|} v / \partial x_1^{q_1} \partial x_2^{q_2}, .. \partial x_d^{q_d}, \quad q = (q_1, q_2, .., q_d), \quad q_k \geq 0, \quad |q| = q_1 + q_2 + .. + q_d\,,$$

$\mathring{H}^1(\Omega) := \{v \in H^1(\Omega) : v|_{\partial\Omega} = 0\}$ is the subspace of functions from $H^1(\Omega)$ vanishing on the boundary $\partial\Omega$.

We use also the abbreviations: a.o. – arithmetic operations, FE - finite element. In the vectors of the space variables $x = (x_1, x_2)$ or $x = (x_1, x_2, x_3)$, sometimes we interchange the positions of variables and write $x = (x_k, x_{3-k})$ and $x = (x_k, x_{k+1}, x_{k+2})$, assuming in the latter case that indices $k + l$ are taking values modulo 3.

2 Poisson equation

2.1 An outline of the approach

In this section, we illustrate basics of the approach of the direct evaluation of equilibrated fields on a simple model problem. Let us consider a boundary value problem for the Poisson equation in the unite square $\Omega = (0,1) \times (0,1)$ with the mixed boundary conditions

$$\begin{gathered} -\Delta u = f(x)\,, \qquad x = (x_1, x_2) \in \Omega\,, \quad \partial\Omega = \overline{\Gamma}_D \cup \overline{\Gamma}_N\,, \\ u|_{\Gamma_D} = 0\,, \quad \partial u/\partial\nu|_{\Gamma_N} = 0\,. \end{gathered} \tag{2.1}$$

where

$$\begin{gathered} \Gamma_D = \{x|\ x_1 \in (0,1]\,, x_2 = 1\} \cup \{(x_1, x_2)|\ x_1 = 1\,, x_2 \in (0,1]\}\ , \\ \Gamma_N = \{x|\ x_1 \in [0,1)\,, x_2 = 0\} \cup \{(x_1, x_2)|\ x_1 = 0\,, x_2 \in [0,1)\}\ , \end{gathered} \tag{2.2}$$

and ν is the distance from the boundary along the outward normal. The generalized formulation of this boundary value problem reads

$$a(u,v) = (f,v)_\Omega\,, \qquad \forall\, v \in \mathbb{V}(\Omega)\,, \tag{2.3}$$

where $\mathbb{V}(\Omega) = \{v \in H^1(\Omega) : v|_{\Gamma_D} = 0\}$

$$a(v,w) = \int_\Omega \nabla v \cdot \nabla w\, dx\,, \qquad (v,w)_\Omega = \int_\Omega v\, w\, dx\,.$$

Let $\mathcal{V}(\Omega)$ be the finite element space of the piece wise bilinear functions on the uniform square grid of size $h = 1/n,\ n > 1,$ with the nodes $x^{(i)} = h(i_1, i_2),\ i_k = 0, 1, .., n,$ and $\mathcal{V}_0(\Omega)$ be the subspace of functions from $\mathcal{V}(\Omega)$, vanishing on Γ_D. By $u_{\rm fem}$ is denoted the finite element solution belonging to $\mathcal{V}_0(\Omega)$ and satisfying the identity

$$a(u_{\rm fem}, \widetilde{v}) = (f, \widetilde{v})_\Omega\,, \qquad \forall\, \widetilde{v} \in \mathcal{V}_0(\Omega)\,. \tag{2.4}$$

In order to be able to efficiently implement the a posteriori estimate (1.1), it is necessary with the use of the obtained FE solution $u_{\rm fem}$ to construct the vector valued function $\mathbf{t} = (t_1, t_2)^\top$, which obeys the two conditions:

α) *it satisfies the balance differential equation*

$$-\nabla \cdot \mathbf{t} = f\,, \tag{2.5}$$

and the boundary conditions

$$\boldsymbol{\nu} \cdot \mathbf{t}|_{\Gamma_N} = 0\,, \tag{2.6}$$

where $\boldsymbol{\nu}$ *is a unite vector normal to* $\partial\Omega$*, and*

β) *is as much close as possible to the gradient* ∇u *of the exact solution.*

We shall use notations Q_f, Q_0 for the sets of functions satisfying (2.5),(2.6) with the given f and $f = 0$, respectively, from which Q_0 is clearly a liner space. The balance law (2.5) models equilibrium equations in the case of the theory elasticity boundary value problems. Elements of Q_f will be termed balanced or equilibrated, whereas elements of Q_0 – self-balanced or self-equilibrated. The sets Q_f, Q_0 can be defined constructively by means of the splitting technique, which was introduced by Korneev and Rozin [45, 29, 27] at developing numerical methods for solving problems of the theory of elastic thin plates and shells in solid mechanics on the basis of the Castigliano principle. For the problem under consideration constructive definition of the set Q_f of the "equilibrated" vectors is quite simple. If q is an arbitrary sufficiently smooth function, then the vector $\mathbf{t} = (t_1, t_2)^\top$ with the components

$$t_1(x) = \int_0^{x_1} q(\xi_1, x_2)\, d\xi_1\,, \qquad t_2(x) = -\int_0^{x_2} (f(x_1, \xi_2) + q(x_1, \xi_2))\, d\xi_2 \tag{2.7}$$

satisfies equation (2.5) and boundary conditions (2.6). Clearly, $Q_f = Q_0 + \mathbf{t}_f$, where $\mathbf{t}_f$ any element of Q_f, and Q_0 is defined by (2.7) with $f \equiv 0$. For all $q \in L_2(\Omega)$ one comes to the space Q_0 with the norm $|\mathbf{t}|_\sigma = ||\mathbf{t}||$, where $\|\cdot\|$ stands for the $[L_2(\Omega)]^2$ norm.

The estimate (1.1) takes the form

$$\|\nabla(u-u_{\rm fem})\| \le \|\nabla u_{\rm fem} - \mathbf{t}\|\,, \qquad \forall\ \mathbf{t}\in Q_f\,. \tag{2.8}$$

Obviously, a better approximation of $\partial u/\partial x_1$ by t_1 (*e.g.,* with the use of values of the gradient of the finite element solution at superconvergence points) will result in a better a posteriori estimate. In turn, from (2.7) it is seen that the function q has the sense of the second derivative $\partial^2 u/\partial x_1^2$.

Taking for $\alpha_k,$ sufficiently smooth functions satisfying $\alpha_1+\alpha_2=1$, one can use more "symmetric" formulas instead of (2.7):

$$\left\{\begin{array}{l} t_1(x) = -\int_0^{x_1}(\alpha_1 f - q)\,(\xi_1,x_2)\,d\xi_1\,,\\[2ex] t_2(x) = -\int_0^{x_2}(\alpha_2 f + q)\,(x_1,\xi_2)\,d\xi_2\,. \end{array}\right. \tag{2.9}$$

They can provide more accurate a posteriori estimates, especially with a good choice of functions $\alpha_k,$, but require more a.o.

If the approach, presented above, is used for a posteriori estimation, then different boundary conditions should be given attention. Suppose, $\widetilde{u}$ is an approximate twice differentiable solution of (2.1),(2.2), *e.g.,* obtained by the Galerkin method or any other function from $\mathbb{V}(\Omega)$. For instance, we can set $q=\partial^2\widetilde{u}/\partial x_1^2$ and come to the expressions

$$\begin{array}{l} t_1 = \int_0^{x_1}\partial^2\widetilde{u}/\partial x_1^2(\xi_1,x_2)d\,\xi_1\,,\\[2ex] t_2 = -\int_0^{x_2}(f+\partial^2\widetilde{u}/\partial x_1^2)(x_1,\xi_2)d\,\xi_2\,. \end{array} \tag{2.10}$$

Similarly, we can proceed from setting $q=\partial^2\widetilde{u}/\partial x_2^2$ and obtain the a posteriori estimates

$$\begin{array}{l} \|\nabla(u-\widetilde{u})\|^2 \le \int_\Omega\Big\{[\frac{\partial\widetilde{u}}{\partial x_k} - \int_0^{x_k}\frac{\partial^2\widetilde{u}}{\partial x_k^2}(\xi_k,x_{3-k})d\,\xi_k]^2 + \\[2ex] +\ [\frac{\partial\widetilde{u}}{\partial x_{3-k}} + \int_0^{x_{3-k}}(f+\frac{\partial^2\widetilde{u}}{\partial x_k^2})(x_k,\xi_{3-k})d\,\xi_{3-k}]^2\Big\}\,d\,x\,, \qquad k=1,2\,. \end{array} \tag{2.11}$$

Taking into account boundary conditions at Γ_N and triangular inequality, one comes from (2.11) to

$$\|\nabla(u-\widetilde{u})\| \le \sum_{k=1,2}\|\frac{\partial\widetilde{u}}{x_k}\big|_{x_k=0}\|_{(0,1)} + \|\int_0^{x_2}(f+\Delta\widetilde{u})(x_1,\xi_2)\,d\,\xi_2\| \tag{2.12}$$

Another vector $\mathbf{t}$, which belongs to Q_f simultaneously with $\mathbf{t}$ from (2.10), is

$$\begin{array}{l} t_1 = \int_0^{x_1}(\alpha_1\partial^2\widetilde{u}/\partial x_1^2 - \alpha_2(f+\partial^2\widetilde{u}/\partial x_2^2))\,(\xi_1,x_2)d\,\xi_1\,,\\[2ex] t_2 = \int_0^{x_2}(\alpha_2\partial^2\widetilde{u}/\partial x_2^2 - \alpha_1(f+\partial^2\widetilde{u}/\partial x_1^2))\,(x_1,\xi_2)d\,\xi_2\,. \end{array} \tag{2.13}$$

The corresponding a posteriori error estimates are

$$\begin{array}{l} \|\nabla(u-\widetilde{u})\| \ \le \Big\{\sum_{k=1,2}\int_\Omega\Big[\frac{\partial\widetilde{u}}{\partial x_k} - \int_0^{x_k}\Big(\alpha_k\frac{\partial^2\widetilde{u}}{\partial x_k^2} - \alpha_{3-k}(f+\frac{\partial^2\widetilde{u}}{\partial x_{3-k}^2})\Big)\,(\xi_k,x_{3-k})d\,\xi_k\Big]^2\,d\,x\Big\}^{1/2} \le \\[2ex] \le \sum_{k=1,2}\Big[\|\frac{\partial\widetilde{u}}{x_k}\big|_{x_k=0}\|_{(0,1)} + \|\int_0^{x_k}\alpha_{3-k}(f+\Delta\widetilde{u})(\xi_k,x_{3-k})\,d\,\xi_k\|\Big] \end{array} \tag{2.14}$$

which for $\alpha_k \equiv 0.5$ is invariant with respect to x_k, $k = 1, 2,$. It easy to see, that adding and subtracting $\alpha_{3-k}\partial^2\widetilde{u}/\partial x_k^2$ inside round brackets, taking into account boundary conditions at Γ_N and triangular inequality, we obtain the same estimate (2.12).

In the case of the Dirichlet boundary value problem

$$a(u,v) = (f,v)_\Omega\,, \qquad u, \forall\, v \in \mathbb{V}(\Omega) = \mathring{H}^1(\Omega)\,, \tag{2.15}$$

the estimates (2.12),(2.14) take especially simple forms. Instead of the latter we have

$$\|\nabla(u-\widetilde{u})\| \le \sum_{k=1,2} \| \int_0^{x_k} \alpha_{3-k}(f+\Delta\widetilde{u})(\xi_k, x_{3-k})\, d\,\xi_k \|\,. \tag{2.16}$$

Let for simplicity $\alpha_2 \equiv 0$. Since in the case under consideration no boundary conditions are imposed on the equilibrated fluxes, we can set

$$\begin{aligned} &t_1 = \partial\widetilde{u}/\partial x_1\,, \\ &t_2 = \partial\widetilde{u}/\partial x_2(x_1,0) - \textstyle\int_0^{x_2}(f+\partial^2\widetilde{u}/\partial x_1^2)(x_1,\xi_2)d\,\xi_2\,, \end{aligned} \tag{2.17}$$

and, therefore,

$$\begin{aligned} \tfrac{\partial\widetilde{u}}{\partial x_1} - t_1 &= 0\,, \\ \tfrac{\partial\widetilde{u}}{\partial x_2} - t_2 &= \tfrac{\partial\widetilde{u}}{\partial x_2} - \tfrac{\partial\widetilde{u}}{\partial x_2}(x_1,0) + \textstyle\int_0^{x_2}(f+\frac{\partial^2\widetilde{u}}{\partial x_1^2})(x_1,\xi_2)d\,\xi_2 = \int_0^{x_2}(f-\Delta\widetilde{u})(x_1,\xi_2)\,d\,\xi_2\,, \end{aligned}$$

completing the proof. In the case of an arbitrary sufficiently smooth domain, the proof is similar.

Lemma 2.1. *Let Ω be Lipshits continuous domain, $f \in L_2(\Omega)$, u - solution of (2.15), and $\widetilde{u}$ be any function in $H^2(\Omega)$ satisfying boundary condition $u\big|_{\partial\Omega} = 0$. Then the error $u - \widetilde{u}$ satisfies a posteriori estimate (2.16).*

2.2 Examples of algorithms for finite element solutions of Poisson equation

Solutions obtained by FE methods compatible in C, which are primarily used in practice for second order elliptic equations, do not have second derivatives. Basically, three ways to outflank this obstacle can be distinguished. All of them start from the procedure of constructing some smooth approximation of the second, or first derivatives of FE solution, or the FE solution itself. In what follows, this procedure is termed *smoothing procedure*. After smoothing procedure has been applied, we proceed in one of the ways described in Subsection 2.1. The distinctions between three types of a posteriori error estimation algorithms for our model problem can be illustrated on example of "nonsymmetric" algorithms, in which 1-d integration of f is involved in the definition of only one of the fluxes. Since this flux is uniquely defined by the balance equation and the boundary condition on Γ_N, it is sufficient to point out the way of evaluation of one flux, which is calculated first. Briefly, three types of such a posteriori error estimation algorithms are the following:

a) Calculate second derivative of the FE solution along one of the axes approximately with the use of finite differences at some set of discrete points (*e.g.,* FE nodes). Define q (*e.g.,* as a function of the FE space $\mathcal{V}(\Omega)$) by interpolation of the calculated approximate values of the second derivative. Define the corresponding flux by 1-d integration of q (like in the first expression (2.7)) and by adding the boundary value of the flux, given by the boundary conditions on Γ_N.

b) Calculate the first derivative of the FE solution in one of the directions x_k at the nodal points, *e.g.* by averaging. Define the tentative flux in the chosen direction as the FE function, which belongs to $\mathcal{V}(\Omega)$ and takes at the nodes calculated values. Define the flux by adjustment of the tentative flux to the boundary condition on Γ_N.

c) Construct twice differentiable approximation of the FE solution. Define the tentative flux in one of the axes as the first derivative of the smoothed FE solution. Define the flux by adjustment of the tentative flux to the boundary condition on Γ_N.

Indeed each of a),b) and c) allows to define two equilibrated fluxes $\mathbf{t}^{(k)},\ k = 1, 2,$ corresponding to the direction x_k, the flux along which is defined first. The flux for the a posteriori estimator can be defined as $\mathbf{t} = \alpha_1\mathbf{t}^{(1)} + \alpha_2\mathbf{t}^{(2)}$.

Apart from "symmetric" versions, other variations of the outlined algorithms are numerous. For instance, in the smoothing procedures of a) and b) some other finite dimensional functional spaces $\mathcal{W}(\Omega)$ can be used instead of the basic FE space $\mathcal{V}(\Omega)$. Since it is not necessary for q to be continuous, we can cover Ω by some nonoverlapping subdomains Ω_j and in each define q by the list squares method with the use of polynomials of some specific order p_j. In the vicinity of singularities in the exact solution of the problem, special representations for q, equilibrated fluxes or the smoothed FE solution can be implemented. In the latter case it is more appropriate to use the term *preprocessed* FE solution. It is also worth adding that in different subdomains one can use algorithms of different types, *i.e.,* a), b) or c), for obtaining equilibrated fluxes.

In this subsection, we present a few simple examples of the outlined algorithms for the Poisson equation in the unit square and arbitrary sufficiently smooth domain with different boundary conditions. For the problems in the unit square, the FE space $\mathcal{V}(\Omega)$ is the space of continuous piece wise bilinear functions on the square mesh of size $h = 1/n$. For the nodes of this mesh we use notation $x^{(i)} = hi,\ i = (i_1, i_2)$. We start from the algorithm of the type c) for the problem (2.1)(2.2).

Algorithm 2.1.

`Step 1.` For each node $x^{(i)} \in \partial\overline{\Omega}$ calculate the value of the mesh function $v_h = (v^{(i)}_{2,1})^n_{i_1,i_2=0}$, which is the finite-difference approximation of the second derivative $\partial^2 u/\partial {x_1}^2(x^{(i)})$. For internal nodes of horizontal mesh lines $x_2 \equiv hi_2,\ i_2 = 0, 1, .., n,$ use

$$v^{(i)}_{2,1} = \frac{u_{\rm fem}(h(i_1+1), hi_2) - 2u_{\rm fem}(hi_1, hi_2) + u_{\rm fem}(h(i_1-1), hi_2)}{h^2}\,, \qquad i_1 = 1, 2.., n-1\,.$$

For the nodes $(0, i_2)$ on the axis $x_1 \equiv 0$ set

$$v^{(i)}_{2,1} = \partial^2\widetilde{u}_0/\partial {x_1}^2(0, hi_2)\,,$$

where $\widetilde{u}_0$ is the 3-rd order interpolation polynomial of x_1 over the values $u_{\rm fem}(x^{(i)})$ for $i_1 = 0, 1, 2,$ and $\partial u_{\rm fem}/\partial x_1(0, hi_2)$. For the nodes (n, i_2) calculate

$$v^{(i)}_{2,1} = \partial^2\widetilde{u}_1/\partial {x_1}^2(1, hi_2)\,,$$

where $\widetilde{u}_1$ is the 3-rd order Lagrange interpolation polynomial of x_1 over the values $u_{\rm fem}(x^{(i)})$ for $i_1 = n-3, n-2, n-1, n$.

`Step 2.` Define $I_{2,1}(x) \subset \mathcal{V}(\Omega)$ as the piece wise bilinear interpolation of v_h.

`Step 3.` Define $\mathbf{t}$ by evaluating the integrals

$$\begin{aligned} t_1(x) &= \textstyle\int_0^{x_1} I_{2,1}(\xi_1, x_2)\, d\xi_1\,, \\ t_2(x) &= -\textstyle\int_0^{x_2} (f(x_1,\xi_2) + I_{2,1}(x_1,\xi_2))\, d\xi_2\,. \end{aligned} \tag{2.18}$$

`Step 4.` Evaluate the estimator $\eta := \|\nabla u_{\rm fem} - \mathbf{t}\|^2$.

Remark 2.1. *Formulas (2.18) correspond to (2.7),(2.10). Since $I_{2,2} \sim \partial^2 u/\partial {x_1}^2$ can be calculated in a similar way, a more "symmetric" formulas, corresponding to (2.9),(2.13), can be used:*

$$\begin{cases} t_1 = \int_0^{x_1} [\alpha_1 I_{2,1} - \alpha_2(f + I_{2,2})](\xi_1, x_2)\, d\xi_1 = \int_0^{x_1} [I_{2,1} - \alpha_2(f + I_{2,1} + I_{2,2})](\xi_1, x_2)\, d\xi_1\,, \\ t_2 = \int_0^{x_2} [\alpha_2 I_{2,2} - \alpha_1(f + I_{2,1})](x_1, \xi_2)\, d\xi_2 = \int_0^{x_2} [I_{2,2} - \alpha_1(f + I_{2,1} + I_{2,2})](x_1, \xi_2)\, d\xi_2\,. \end{cases} \tag{2.19}$$

In `Step 1`, we used finite-difference approximations of the second derivative $\partial^2 u/\partial x_1^2$, which provide the same order h^2 of approximation for a sufficiently smooth u. In `Step 3`, the way of the evaluation of integrals of I_v and f may be different. In particular, in many cases the analytical integration can be performed. In general, procedures, used in `Steps 3`, depend on the way of evaluating the norm $\|\nabla u_{\rm fem} - \mathbf{t}\|^2$ in `Step 4`. For instance, for each finite element $\tau_i := h(i_1 - 1, i_1) \times h(i_2 - 1, i_2)$, we can use quadratures exact for polynomials of some order $p_i \geq 1$. Then it is necessary to evaluate t_k only at the quadrature nodes, and for doing this other quadratures can be used. The type of the quadratures, used element wise for evaluating t_2 and the norm $\|\nabla u_{\rm fem} - \mathbf{t}\|^2$ may depend on the local smoothness of f, if the integral of f is not evaluated exactly. Bounds for quadrature errors will enter the resulting error bound for the FE solution with the right part depending only on $u_{\rm fem}$ and f. However, we will not elaborate on these subjects in the present paper aimed to illustrate main features of the approach.

In the case $\Gamma_N = \emptyset$, *i.e.* when only the Dirichlet boundary conditions are imposed in (2.1), then the estimate (2.8) is true for any vector valued function $\mathbf{t}$ satisfying the balance equation (2.5) and not subjected any boundary conditions.. But when boundary conditions are different from the ones considered above in (2.1) and $\Gamma_N \neq \emptyset$ then some remedy should be done in order $\mathbf{t}$ to fulfill the boundary conditions (2.6). For instance, let

$$\Gamma_D = \{(x_1, x_2) |\ x_1 \in [0,1]\,, x_2 = 1\}\,, \qquad \Gamma_N = \Gamma_{N,1} \cup \Gamma_{N,2} \cup \Gamma_{N,3}\,,$$

where

$$\begin{aligned} \Gamma_{N,1} &= \{(x_1, x_2) |\ x_2 \in [0,1]\,, x_1 = 0\}\,, \\ \Gamma_{N,2} &= \{(x_1, x_2) |\ x_1 \in [0,1]\,, x_2 = 0\}\,, \\ \Gamma_{N,3} &= \{(x_1, x_2) |\ x_2 \in [0,1]\,, x_1 = 1\}\,, \end{aligned}$$

and the boundary conditions are

$$u|_{\Gamma_D} = 0\,, \quad \partial u/\partial \nu|_{\Gamma_N} = g\,, \tag{2.20}$$

where $g \in C(\overline{\Gamma}_N)$ and

$$g = \begin{cases} g_1(x_2), & \text{on } \Gamma_{N,1}\,, \\ g_2(x_1), & \text{on } \Gamma_{N,2}\,, \\ g_3(x_2), & \text{on } \Gamma_{N,3}\,. \end{cases} \tag{2.21}$$

The estimate (2.8) is true, if the vector valued function $\mathbf{t}$ satisfies the equation (2.5) and the boundary condition (2.6). The steps 1 and 2 may be the same as in the algorithm adduced above, but they produce only a tentative flux, which should be adjusted to the boundary condition (2.6). Let

$$\begin{aligned} t_1(x) &= \int_0^{x_1} (I_{2,1} + r)(\xi_1, x_2)\, d\xi_1 \ + g_1(x_2)\,, \\ t_2(x) &= -\int_0^{x_2} (f + I_{2,1} + r)(x_1, \xi_2)\, d\xi_2 + g_2(x_1)\,, \end{aligned} \tag{2.22}$$

where $r(x)$ is chosen with the purpose to fulfill the boundary condition (2.6) on $\Gamma_{N,3}$. It is sufficient to take r depending only on x_2

$$r(x_1, x_2) = g_3(x_2) - g_1(x_2) - \int_0^1 I_{2,1}(\xi_1, x_2)\, d\xi_1\,. \tag{2.23}$$

Hence

$$\begin{aligned} t_1(x) &= \int_0^{x_1} I_{2,1}(\xi_1, x_2)\, d\xi_1 - x_1 \int_0^1 I_{2,1}(\xi_1, x_2)\, d\xi_1 + (1 - x_1) g_1(x_2) + x_1 g_3(x_2)\,, \\ t_2(x) &= -\int_0^{x_2} (f + I_{2,1})(x_1, \xi_2)\, d\xi_2 + \int_0^{x_2} \int_0^1 I_{2,1}(\xi_1, \xi_2))\, d\xi_1 d\xi_2 + \\ &+ \int_0^{x_2} (g_1 - g_3)(\xi_2)\, d\xi_2 + g_2(x_1)\,. \end{aligned} \tag{2.24}$$

In the case of the homogeneous Neumann boundary condition on Γ_N, the expressions (2.24) take a simpler form

$$\begin{aligned} t_1 &= \int_0^{x_1} I_{2,1}(\xi_1, x_2)\, d\xi_1 - x_1 \int_0^1 I_{2,1}(\xi_1, x_2)\, d\xi_1\,, \\ t_2 &= -\int_0^{x_2} (f + I_{2,1})(x_1, \xi_2)\, d\xi_2 + \int_0^{x_2} \int_0^1 I_{2,1}(\xi_1, \xi_2))\, d\xi_1 d\xi_2\,. \end{aligned} \tag{2.25}$$

The described approach is easily realized in a much more general situation. Suppose, Ω is the domain occupied by the arbitrary triangulation $\mathbb{S}_h$ with the triangles $\tau_r,\ r = 1, 2, .., \mathcal{R},\ \mathcal{V}(\Omega)$ is the space of the continuous piece wise linear functions and $\mathring{\mathcal{V}}(\Omega) = \{v \in \mathcal{V}(\Omega) : v\big|_{\partial\Omega=0}\}$. We will turn to the algorithm of the type b) for the problem with the Dirichlet boundary condition. Namely, we consider the problem

$$a(u, v) = (f, v)_\Omega\,, \qquad u, \forall\, v \in \mathbb{V}(\Omega) = \mathring{H}^1(\Omega)\,, \tag{2.26}$$

and its FE solution u_{fem} satisfying

$$a(u_{\text{fem}}, v) = (f, v)_\Omega\,, \qquad u_{\text{fem}}, \forall\, v \in \mathring{\mathcal{V}}(\Omega)\,. \tag{2.27}$$

Let us assume for simplicity that each line $x_k \equiv \text{const}$ crosses Ω not more than in two points, $\Gamma_{k,-}$ is the part of the boundary containing the points of such pairs, having lesser coordinate x_k, $x_k = a_k(x_{3-k})$ is the equation of $\Gamma_{k,-}$, $x_{3-k} = \hat{a}_{3-k}, \hat{b}_{3-k}$ are the coordinates of the ends of $\Gamma_{k,-}$. We also use the notation $B_i = \{r : \tau_r \in \Omega,\, x^{(i)} \in \overline{\tau}_r\}$ for the set of the numbers of finite elements, having $x^{(i)}$ for a vertex, and notation $\mathcal{V}(\Gamma_{k,-})$ for the space of the traces of functions from $\mathcal{V}(\Omega)$ on $\Gamma_{k,-}$. Among simplest is the algorithm, which is not invariant with respect to x_k, $k = 1, 2$, and is based on the averaging and the procedure reflected in Remark 2.1.

Algorithm 2.2.

`Step 1.` For each node calculate the average

$$\overline{u}^{(i)}_{\text{fem},1} = \sum_{r \in B_i} \frac{\partial u^{(r)}_{\text{fem}}}{\partial x_1}(x^{(i)})\,,$$

where $u^{(r)}_{\text{fem}} = u_{\text{fem}}\Big|_{\tau_r}$ is the restriction of the FE solution to τ_r.

`Step 2.` Define the interpolation $I(\overline{u}_{\text{fem},1}) \in \mathcal{V}(\Omega)$ satisfying

$$I(\overline{u}_{\text{fem},1})(x^{(i)}) = \overline{u}^{(i)}_{\text{fem},1}\,, \qquad \forall\ x^{(i)} \in \overline{\Omega}\,.$$

`Step 3.` For each node $x^{(i)} \in \overline{\Gamma}_{2,-}$ calculate the average

$$\overline{u}^{(i)}_{\text{fem},2} = \sum_{r \in B_i} \frac{\partial u^{(r)}_{\text{fem}}}{\partial x_2}(x^{(i)})\,,$$

and for $x \in \overline{\Gamma}_{2,-}$ define the piece wise linear continuous interpolation $I_{\Gamma_{k,-}}(\overline{u}_{\text{fem},2})$ satisfying

$$I(\overline{u}_{\text{fem},2})(x^{(i)}) = \overline{u}^{(i)}_{\text{fem},2}\,, \qquad \forall\ x^{(i)} \in \overline{\Gamma}_{2,-}\,.$$

`Step 4.` For components of an equilibrated vector $\mathbf{t} = (t_1, t_2)^\top$ set

$$t_1 = I(\overline{u}_{\text{fem},1})\,, \qquad t_2\big|_{\Gamma_{2,-}} = t_2(x_1, \phi_2(x_1)) := I_{\Gamma_{k,-}}(\overline{u}_{\text{fem},2})\big|_{\Gamma_{2,-}}\,,$$

and evaluate

$$t_2(x) = t_2(x_1, \phi_2(x_1)) - \int_{a_2}^{x_2} (f + \frac{\partial t_1}{\partial x_1})(x_1, \xi_2)\, d\xi_2\,.$$

`Step 5.` Evaluate the bound $\|\nabla u_{\text{fem}} - \mathbf{t}\|^2$.

The algorithm, based on averaging, can be made invariant with respect to x_k, $k = 1, 2,$.

Algorithm 2.3.

`Step 1.` For each node and $k = 1, 2$ calculate the averages

$$\overline{u}^{(i)}_{\text{fem},k} = \sum_{r \in B_i} \frac{\partial u^{(r)}_{\text{fem}}}{\partial x_k}(x^{(i)})\,.$$

`Step 2.` Define the interpolations $I(\overline{u}_{\text{fem},k}) \in \mathcal{V}(\Omega)$ satisfying

$$I(\overline{u}_{\text{fem},k})(x^{(i)}) = \overline{u}^{(i)}_{\text{fem},k}\,, \qquad \forall\ x^{(i)} \in \overline{\Omega}\,.$$

Clearly, the vector $\widetilde{\mathbf{t}} = (\widetilde{t}_1, \widetilde{t}_2)^\top$, $\widetilde{t}_k = I(\overline{u}_{\mathrm{fem},k})$, approximates the equilibrated vector.

`Step 3.` For $k = 1, 2$ calculate

$$q_{\mathrm{fem},k} = -\partial I(\overline{u}_{\mathrm{fem},k})/\partial x_k\,, \qquad \delta f = f - q_{\mathrm{fem},1} - q_{\mathrm{fem},2}\,.$$

`Step 4.` For $\alpha_1(x) + \alpha_2(x) \equiv 1$ and $\theta_k(x_{3-k}) = I(\overline{u}_{\mathrm{fem},k})(a_k(x_{3-k}), x_{3-k})$, $k = 1, 2$, calculate

$$\begin{aligned} t_k(x) \;&= \theta_k(x_{3-k}) - \textstyle\int_{a_k}^{x_k} (q_{\mathrm{fem},k} + \alpha_k \delta f)(\xi_k, x_{3-k})\, d\xi_k = \\ &= (1-\alpha_k) I(\overline{u}_{\mathrm{fem},k}) - \textstyle\int_{a_k}^{x_k} \alpha_k (f + q_{\mathrm{fem},3-k})(\xi_k, x_{3-k})\, d\xi_k\,. \end{aligned}$$

`Step 5.` Evaluate the bound $\|\nabla u_{\mathrm{fem}} - \mathbf{t}\|^2$.

Functions $\theta_k(x_{3-k})$ specify boundary conditions for $\mathbf{t}$. They can be defined differently from `Step 4`, and for their evaluation by means of the FE solution special more accurate procedures can be used. The simplest choice for α is $\alpha \equiv 1/2$, however a number of more sophisticated procedures for the evaluation of this function can be considered. For instance, in some regions α can be chosen on the basis of the local analysis. Obviously, the procedures of Algorithm 2.2 are a part of Algorithm 2.3. The whole process of the a posteriori estimation can be arranged in the following way. One uses Algorithm 2.2. If the a posteriori estimate is unsatisfactory then additional calculations are performed according Algorithm 2.3 with some choice of α_k. Further perfection of the a posteriori estimator is possible in a variety of ways. For instance, it is not necessary to use the same FE mesh for the evaluation of $\mathbf{t}$. In order to simplify the computations and make the estimate more accurate, a subsidiary mesh can be used for finding $\mathbf{t}$, which, *e.g.,* is orthogonal inside the domain and provide some *hp* FE interpolation for fluxes, determined by the FE solution u_{fem}.

Remark 2.2. *Evaluation of a posteriori error bounds according Algorithm 2.1-Algorithm 2.3 involves only three operations*

`-- numerical differentiation with the use of finite differences,`
`-- interpolation, and`
`-- evaluating of 1-d integrals`.

For this reason, these algorithms are obviously optimal in the arithmetic operations count, if the mesh is orthogonal. In the case of an arbitrary quasiuniform triangulation, the evaluation of integrals may be often arranged by layers of elements. From layer to layer the number of points, at which we need to evaluate an equilibrated flux may in general double. Therefore, the computational cost of the third among listed operations is estimated as $\mathcal{O}(n_k n_{3-k}^2)$, where n_k is the maximal number of nodes in one layer and n_{3-k} is the number of layers. In this paper, we concentrate on basic facts of a posteriori estimation, but several recipes can be immediately suggested for the reduction of the computational work, even in the case of nonuniform unstructured meshes. For instance, we can cover the computational domain by the nonuniform orthogonal mesh with the hanging nodes, matching in density the FE grid. Then we calculate one or both fluxes at the nodes of this mesh by means of averaging and interpolation. After that with the use of the introduced orthogonal mesh, we obtain corrections, which are necessary in order to make fluxes equilibrated.

Remark 2.3. *There are known a posteriori estimates for the Dirichlet problem (2.26)*

$$\begin{aligned}
&\|\nabla u_{\rm fem} - \nabla u\|^2_{0,\Omega} \le (1+\epsilon)\|\nabla u_{\rm fem} - \mathbf{y}\|^2_{0,\Omega} + (1+\tfrac{1}{\epsilon})\|\nabla\cdot\mathbf{y} - f\|^2_{-1,\Omega}\,,\\
&\|\nabla u_{\rm fem} - \nabla u\|^2_{0,\Omega} \le (1+\epsilon)\|\nabla u_{\rm fem} - \mathbf{y}\|^2_{0,\Omega} + c_\Omega(1+\tfrac{1}{\epsilon})\|\nabla\cdot\mathbf{y} - f\|^2_{0,\Omega}\,, \qquad \forall\ \epsilon>0\,,
\end{aligned} \tag{2.28}$$

where $\mathbf{y}$ *is an arbitrary vector and* c_Ω *is the constant from the Friedrich's inequality, see Repin/Frolov [41] and Neittaanmäkki/Repin [35]. One of our estimates can be written in the form*

$$\begin{aligned}
&\|\nabla u_{\rm fem} - \nabla u\|^2_{0,\Omega} \le \|\nabla u_{\rm fem} - \mathbf{t}\|^2_{0,\Omega} \le\\
&\le (1+\epsilon)\|\nabla u_{\rm fem} - \mathbf{y}\|^2_{0,\Omega} + (1+\tfrac{1}{\epsilon})\|\textstyle\sum_{k=1,2}\int_{a_k(x_{3-k})}^{x_k}\alpha_k(f-\nabla\cdot\mathbf{y})(\eta_k,x_{3-k})\,d\eta_k\|^2_{0,\Omega}\,.
\end{aligned} \tag{2.29}$$

In particular, in Algorithm 2.3 we used $\mathbf{y} = (I(\overline{u}_{\rm fem,1}), I(\overline{u}_{\rm fem,2}))^\top$. *The right part* f *enters the last a posteriori estimate in a more adequate way, than in the second estimate (2.28). At the same time, it is easily computable, whereas the negative norm, entering the right part of the first bound (2.28), makes the bound difficult for the use in practice.*

2.3 Heat conduction problem with discontinuous coefficient

The extension of the advocated approach and in particular of the algorithms of the previous section to the elliptic equations with discontinuous coefficients is straightforward. Let us consider as an example the boundary value problem

$$\begin{aligned}
-\quad &\nabla\cdot(\rho(x)\nabla u) = f(x)\,, \qquad x\in\Omega=(0,1)\times(0,1)\,,\\
&u\big|_{\Gamma_D} = g\,, \quad \tfrac{\partial u}{\partial\nu}\big|_{\Gamma_N} = 0\,,
\end{aligned} \tag{2.30}$$

with Γ_D, Γ_N defined as in (2.2), $\rho(x)>0$ and

$$\rho(x) = \begin{cases} \rho_1 = \text{const} & \text{for} \quad x\in\Omega_1 := \{x\in\Omega : 0<x_1<0.5\}\,,\\ \rho_2 = \text{const} & \text{for} \quad x\in\Omega_2 := \Omega\setminus\overline{\Omega}_1\}\,. \end{cases}$$

For simplicity it is assumed that the boundary conditions are consistent and there exits such $u_0\in H^2(\Omega)$ that $u_0\big|_{\Gamma_D} = g$. We define the approximate solution $u_{\rm fem}$ of this problem as a function belonging to the set $\mathcal{L}(\Omega) = \mathcal{V}_0(\Omega) + u_0$ and satisfying the identity

$$a_\rho(u_{\rm fem}, \widetilde{v}) = (f,\widetilde{v})_\Omega\,, \qquad \forall\,\widetilde{v}\in\mathcal{V}_0(\Omega)\,, \tag{2.31}$$

where

$$a_\rho(v,w) = \int_\Omega \rho\nabla v\cdot\nabla w\,dx_1dx_2$$

and $\mathcal{V}_0(\Omega)$ is the FE space defined in Subsection 2.1. If to introduce the norms

$$\|v\|_\rho = (a_\rho(v,v))^{1/2}\,, \qquad \|\mathbf{t}\|_{\rho^{-1}} = (\int_\Omega \rho^{-1}\mathbf{t}\cdot\mathbf{t}\,dx)^{1/2} \tag{2.32}$$

and imply by Q_f the set of the equilibrated vectors satisfying (2.5),(2.6), then the a posteriori error estimate (2.8) takes the form

$$\|\nabla(u-u_{\rm fem})\|_\rho \le \|\rho\nabla u_{\rm fem}-\mathbf{t}\|_{\rho^{-1}}\,. \tag{2.33}$$

Algorithms 2.1-2.3 are easily adapted to the problem under consideration. For instance, the first one is written as follows.

Algorithm 2.4.

`Step 1.` For each node $x^{(i)}\in\partial\overline{\Omega}$ calculate the value of the mesh function $v_h=(v_{2,1}(i))_{i_1,i_2=0}^n$, which is the finite-difference approximation of $\partial(\rho\partial u/\partial x_1)\partial x_1(x^{(i)})$. For internal nodes of horizontal mesh lines $x_2\equiv hi_2,\ i_2=0,1,..,n,$ use

$$v_{1,1}(i)=\rho((i_1-0.5)h,i_2)[u_{\rm fem}(hi)-u_{\rm fem}(h(i_1-1),hi_2)]\,,$$

$$v_{2,1}(i)=\tfrac{1}{h^2}[v_{1,1}(i_1+1,i_2)-v_{1,1}(i)]\,,\qquad i_1=1,2..,n-1\,.$$

For the nodes $(0,i_2)$ on the axis $x_1\equiv 0$ set

$$v_{2,1}(i)=\rho_1\partial^2\widetilde{u}_0/\partial {x_1}^2(0,hi_2)\,,$$

where $\widetilde{u}_0$ is the 3-rd order interpolation polynomial of x_1 over the values $u_{\rm fem}(x^{(i)})$ for $i_1=0,1,2,$ and $\partial u_{\rm fem}/\partial x_1(0,hi_2)$. For the nodes (n,i_2) calculate

$$v_{2,1}(i)=\rho_2\partial^2\widetilde{u}_1/\partial {x_1}^2(1,hi_2)\,,$$

where $\widetilde{u}_1$ is the 3-rd order Lagrange interpolation polynomial of x_1 over the values $u_{\rm fem}(x^{(i)})$ for $i_1=n-3,n-2,n-1,n$.

`Step 2.` Define $I_{2,1}(x)\subset\mathcal{V}(\Omega)$ as the piece wise bilinear interpolation of v_h.

`Step 3.` Define $\mathbf{t}$ by evaluating the integrals

$$\begin{aligned} t_1(x)&=\textstyle\int_0^{x_1} I_{2,1}(\xi_1,x_2)\,d\xi_1\,,\\ t_2(x)&=-\textstyle\int_0^{x_2}(f(x_1,\xi_2)+I_{2,1}(x_1,\xi_2))\,d\xi_2\,. \end{aligned} \tag{2.34}$$

`Step 4.` Evaluate $\|\rho\nabla u_{\rm fem}-\mathbf{t}\|^2_{\rho^{-1}}$.

3 Linear elasticity problems

3.1 A posteriori estimation for plane problems

Let E be the Young's modulus, ν – the Poisson's ratio, $\mathbf{I}$ – the unit tensor and $\mathrm{tr}(\boldsymbol{\kappa})=\boldsymbol{\kappa}:\mathbf{I}$ – the trace of a tensor $\boldsymbol{\kappa}$. The linearly elastic plain strain problem in some domain Ω is formulated in terms of the displacement vector $\mathbf{u}(x)=(u_1(x),u_2(x))^\top$ and symmetric strain and stress tensors

$$\boldsymbol{\sigma}=\begin{pmatrix}\sigma_{11}&\sigma_{12}\\ \sigma_{21}&\sigma_{22}\end{pmatrix}\,,\qquad \boldsymbol{\varepsilon}=\begin{pmatrix}\varepsilon_{11}&\varepsilon_{12}\\ \varepsilon_{21}&\varepsilon_{22}\end{pmatrix}\,,$$

related by the system of equations

$$\operatorname{div}\boldsymbol{\sigma} + \mathbf{f} = 0\,, \tag{3.1}$$

$$\boldsymbol{\varepsilon}(\mathbf{u}) = (\varepsilon_{kl}(\mathbf{u}))_{k,l=1,2}\,, \qquad \varepsilon_{kl} = \frac{1}{2}(\partial u_k/\partial x_l + \partial u_l/\partial x_k)\,, \tag{3.2}$$

$$\boldsymbol{\sigma}(\boldsymbol{\varepsilon}) = \frac{E}{1+\nu}\left[\frac{\nu}{1-2\nu}\operatorname{tr}(\boldsymbol{\varepsilon})\,\mathbf{I} + \boldsymbol{\varepsilon}\right]\,, \tag{3.3}$$

supplied by boundary conditions. The mixed homogeneous boundary conditions may have the form

$$\mathbf{u}|_{\Gamma_D} = 0\,, \qquad \boldsymbol{\sigma}_n = \boldsymbol{\tau}_n|_{\Gamma_N} = 0\,, \tag{3.4}$$

where $\boldsymbol{\sigma}_n(x)$ and $\boldsymbol{\tau}_n(x)$ are stresses normal and tangential to the boundary at a point $x \in \Gamma_N$.

We turn to the case of the first boundary value problem when $\Gamma_N = \varnothing$ and for the boundary conditions we have

$$\mathbf{u}|_{\partial\Omega} = 0\,. \tag{3.5}$$

We assume that for positive constants $\underline{c}_E, .., \overline{c}_\nu$

$$\underline{c}_E \le E(x) \le \overline{c}_E\,, \qquad \underline{c}_\nu \le \nu(x) \le \overline{c}_\nu < 0.5\,,$$

and introduce the space $\mathbb{V} = [\mathring{H}^1(\Omega)]^2$. The generalized solution $\mathbf{u}$ of the problem (3.1)–(3.3),(3.5) formulated in respect to displacements satisfies

$$\int_\Omega \boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v})\,dx = \int_\Omega \mathbf{f}\cdot\mathbf{v}\,dx\,, \qquad \forall\,\mathbf{v} \in \mathbb{V}\,. \tag{3.6}$$

For simplicity it is assumed that the domain of the FE assemblage coincides with Ω. For approximate solution of (3.6), we use the subspace $\mathring{\boldsymbol{\mathcal{V}}} = [\mathring{\mathcal{V}}(\Omega)]^2 = [\mathcal{V}(\Omega)]^2 \cap \mathbb{V}$, where $\mathcal{V}(\Omega)$ is the space of FE scalar functions. The FE solution is found from the identity

$$\int_\Omega \boldsymbol{\sigma}(\mathbf{u}_{\rm fem}) : \boldsymbol{\varepsilon}(\mathbf{v})\,dx = \int_\Omega \mathbf{f}\cdot\mathbf{v}\,dx\,, \qquad \forall\,\mathbf{v} \in \mathring{\boldsymbol{\mathcal{V}}}(\Omega)\,. \tag{3.7}$$

The Hook's law (3.3) can be written in the inversed form

$$\boldsymbol{\varepsilon}(\boldsymbol{\sigma}) = \frac{1+\nu}{E}[\boldsymbol{\sigma} + \nu\operatorname{tr}(\boldsymbol{\sigma})\,\mathbf{I}]\,, \tag{3.8}$$

and the both relations define the norms for arbitrary tensors $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$

$$\|\boldsymbol{\varepsilon}\|_\varepsilon = \left(\int_\Omega \boldsymbol{\sigma}(\boldsymbol{\varepsilon}) : \boldsymbol{\varepsilon}\,dx\right)^{1/2}, \qquad \|\boldsymbol{\sigma}\|_\sigma = \left(\int_\Omega \boldsymbol{\sigma} : \boldsymbol{\varepsilon}(\boldsymbol{\sigma})\,dx\right)^{1/2}. \tag{3.9}$$

If $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$ satisfy (3.8), then clearly $\|\boldsymbol{\varepsilon}\|_\varepsilon = \|\boldsymbol{\sigma}\|_\sigma$, and, if additionally $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\mathbf{u})$ and $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(\mathbf{u})$ are related by equations (3.2),(3.3), then the energy norm for displacements is defined according to the expression

$$\|\mathbf{u}\|_U = \left(\int_\Omega \boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{u})\,dx\right)^{1/2} \tag{3.10}$$

The set $\mathbf{Q}_f$ of equilibrated stress tensors is specified as $\mathbf{Q}_f = \boldsymbol{\tau} + \mathbf{Q}_0$, where $\boldsymbol{\tau}$ is any tensor satisfying the equilibrium equations

$$\operatorname{div} \boldsymbol{\tau} + \mathbf{f} = 0 \tag{3.11}$$

and $\mathbf{Q}_0$ is the linear space of *self-equilibrated* functions, *i.e.,* satisfying equilibrium equations with $\mathbf{f} \equiv \mathbf{0}$. The latter space can be considered as a Hilbert space with the scalar product

$$[\boldsymbol{\sigma}, \boldsymbol{\sigma}']_\sigma = \int_\Omega \boldsymbol{\sigma} : \boldsymbol{\varepsilon}(\boldsymbol{\sigma}')\, dx$$

and the norm $\|\cdot\|_\sigma$. For the exact and the FE solutions $\mathbf{u} = (u_1, u_2)^\top$, $\mathbf{u}_{\rm fem} = (u_{{\rm fem},1}, u_{{\rm fem},2})^\top$, respectively, of the plain strain problem (3.6), we have the a posteriori estimate

$$\|\mathbf{u} - \mathbf{u}_{\rm fem}\|_U \le \|\boldsymbol{\sigma}(\mathbf{u}_{\rm fem}) - \boldsymbol{\tau}\|_\sigma\,, \qquad \forall\ \boldsymbol{\tau} \in \mathbf{Q}_f\,. \tag{3.12}$$

For obtaining a good tensor $\boldsymbol{\tau}$, we use the same approach as before. The system of the equilibrium equations is underdetermined and specify the set $\mathbf{Q}_f$ of equilibrated vectors up to the linear space $\mathbf{Q}_0$ of the *self-equilibrated* functions, *i.e.,* satisfying equilibrium equations with $\mathbf{f} \equiv \mathbf{0}$. This fact allows a simple way of a constructive definition of the whole set $\mathbf{Q}_f$, the essence of which is the following: we specify one of the components τ_{kl} by a sufficiently smooth arbitrary function and find other components from the equilibrium equations. Note, where no special assumptions on $\mathbf{f} \in \mathbf{L}_2(\Omega)$ are made, we always assume that $\mathbf{f} \in \mathbf{L}_2(\Omega)$. In algorithms of a posteriori estimators, it is sufficient to point out the way of definition of sufficiently smooth tensors $\boldsymbol{\tau} \in \mathbf{Q}_f$.

Algorithm A.

1. We specify an arbitrary function $\psi_{12} \in L^1_\infty(\Omega)$, arbitrary functions $\psi_{kk,\Gamma_{k,-}}(x_{3-k}) \in L^1_\infty[\hat a_{3-k}, \hat b_{3-k}]$, and set $\tau_{12} = \psi_{12}$.

2. Find $q_1 = \partial\psi_{12}/\partial x_2$ and

$$\tau_{11} = \psi_{11,\Gamma_{1,-}}(x_2) - \int_{a_1(x_2)}^{x_1} (f_1 + q_1)(\xi_1, x_2)\, d\xi_1\,. \tag{3.13}$$

3. Find $q_2 = \partial\psi_{12}/\partial x_1$ and

$$\tau_{22} = \psi_{22,\Gamma_{2,-}}(x_1) - \int_{a_2(x_1)}^{x_2} (f_2 + q_2)(x_1, \xi_2)\, d\xi_2\,. \tag{3.14}$$

As well one can also start from specifying

- two functions $\psi_{kk,\Gamma_{k,-}}(x_{3-k}) \in L_\infty[\hat a_{3-k}, \hat b_{3-k}]$,
- function $\psi_{12,\Gamma_{k,-}}(x_{3-k}) \in L_\infty[\hat a_{3-k}, \hat b_{3-k}]$, and
- function $q(x)$.

Then the tangential stress is defined by the integral

$$\tau_{12} = \psi_{12,\Gamma_{k,-}}(x_{3-k}) - \int_{a_k(x_{x_{3-k}})}^{x_k} q(\xi_k, x_{3-k})\, d\xi_k$$

for one of $k = 1, 2$, and other stresses are defined according to the above algorithm.

If steps 1-3 are used for obtaining an a posteriori estimate, functions ψ_{12} and $\psi_{kk,\Gamma_{k,-}}$ are calculated by means of the obtained FE solution. This should be done in the most accurate way (*e.g.,* with the use of superconvergence properties of the FE solution), since the closeness of these functions to the true stresses $\sigma_{12}(\mathbf{u})$ on Ω and to $\sigma_{kk}(\mathbf{u})$ on the part of the boundary $\Gamma_{k,-}$ is crucial for the accuracy of the a posteriori estimate. In particular, ψ_{12} can be specified as an element of the FE space $\mathcal{V}(\Omega)$ with the nodal values obtained by the procedure of averaging similar to the used in `Steps 2,3` of Algorithm 2.2. In the case of a rectangular domain and orthogonal mesh, this procedure is especially simple

$$\begin{aligned}\psi_{12}(x^{(\mathbf{i})}) &= \sigma_{\mathrm{fem},12}(i_1h-0,i_2h-0)+\sigma_{\mathrm{fem},12}(i_1h+0,i_2h-0)+\\ &+ \sigma_{\mathrm{fem},12}(i_1h-0,i_2h+0)+\sigma_{\mathrm{fem},12}(i_1h+0,i_2h+0)\,.\end{aligned} \tag{3.15}$$

Here $\sigma_{\mathrm{fem},12}=\sigma_{12}(\mathbf{u}_{\mathrm{fem}})$ is the stress defined by the FE solution and it is assumed that $\sigma_{\mathrm{fem},12}(i_1h\pm 0,i_2h\pm 0)=0$, if the corresponding element τ_j, $j=(j_1\pm 1,j_2\pm 1)$ does not belong to Ω.

In the case of a more general domain, $\Gamma_N\subset(\Gamma_{1,-}\cap\Gamma_{2,-})$ and the nonhomegeneous boundary condition on Γ_N, *i.e.,*

$$\boldsymbol{\sigma}_n\big|_{\Gamma_N}=\mathbf{t}\,, \tag{3.16}$$

functions $\psi_{kl,\Gamma_{k,-}}(x_{3-k})$, $x_{3-k}\in(\hat{a}_{3-k},\hat{b}_{3-k})$, on the subset corresponding to Γ_N should satisfy (3.16). Vector $\mathbf{t}$ can be replaced by some more convenient for the use in (3.13),(3.14), but sufficiently accurate approximation $\widetilde{\mathbf{t}}$. The estimate of the error $\mathbf{t}-\widetilde{\mathbf{t}}$ of approximation in an appropriate norm will enter the right part of the a posteriori estimate.

The approach under consideration may be realized in a number of ways. For the first step one can specify $\tau_{kk}=\psi_{kk}$ and then find $\tau_{12},\tau_{3-k,3-k}$ from the equilibrium equations. However, such a path is not invariant with respect to $x_k,k=1,2$. Besides, in the process of obtaining such equilibrated symmetric stress tensor, this function is differentiated twice in x_k and integrated twice along x_{3-k}. The latter requires more smoothness from ψ_{kk} at least in x_k. In general an additional differentiation will certainly result in cruder a posteriori error estimates, if it is not compensated by the integration along the same direction, and this was observed in our numerical experiments. However, more complicated algorithms, but having the same asymptotical computational complexity, may be designed, which are invariant with respect to $x_k,k=1,2$, and at the same time provide continuous equilibrated stress tensors. An example of such algorithms for the problem (3.1)–(3.3),(3.5) in $\Omega=(0,1)\times(0,1)$ is Algorithm B.

Algorithm B.

1. Using the finite element solution, define a continuous piece wise bilinear function $I(\varkappa^{\,h}_{12})\in\mathring{\mathcal{V}}$, which "approximates" the second mixed derivative

$$I(\varkappa^{\,h}_{12})\sim\frac{\partial\sigma_{12}}{\partial x_1\partial x_2}=\frac{E}{2(1+\nu)}\Big(\frac{\partial^3u_1}{\partial x_1\partial x_2^2}+\frac{\partial^3u_2}{\partial x_1^2\partial x_2}\Big)$$

of the stress $\sigma_{12}=\sigma_{12}(\mathbf{u})$. Here, $\varkappa^{\,h}_{12}=\{\varkappa_{12}(i)\}$ is the mesh function and, *e.g.,* $I(\varkappa^{\,h}_{12})(x^{(i)})=\varkappa_{12}(i)$. Inverted commas stand for the reason that indeed $\partial\sigma_{12}(\mathbf{u}_{\mathrm{fem}})/\partial x_1\partial x_2$ may not be defined even on finite elements, and, therefore, some special ways of evaluation of $\varkappa^{\,h}_{12}$ should be implemented. They should allow us to expect approximation in some sense of $\sigma_{12}(\mathbf{u})$ by $\tau_{12}(\mathbf{u}_{\mathrm{fem}})$, evaluated in a posteriori estimator by means of $\varkappa^{\,h}_{12}$. For instance, below τ_{12} is defined by the

backward double integration of $I\left(\varkappa_{12}^{h}\right)$ in such a way that under some conditions one can expect the same accuracy from τ_{12} as from $\sigma_{12}(\mathbf{u}_{\rm fem})$. If the square bilinear or higher order elements are used, one can evaluate $\partial\sigma_{12}(\mathbf{u}_{\rm fem})(x^{(i)})/\partial x_1\partial x_2$ for each element at its nodes, than for each node of FE assemblage calculate $\varkappa_{12}(i)$ as the average of $\partial\sigma_{12}(\mathbf{u}_{\rm fem})(x^{(i)})/\partial x_1\partial x_2$ for each element.

2. Evaluate

$$\begin{aligned}
\tau_{12} &= \textstyle\int_0^{x_2}\int_0^{x_1} I\left(\varkappa_{12}^{h}\right)\,dx_1\,dx_2 + c_0 + \int_0^{x_1} c_1(x_1)\,dx_1 + \int_0^{x_2} c_2(x_2)\,dx_2\,,\\
\tau_{11} &= \textstyle\int_0^{x_1}\left[f_1 - \int_0^{x_1} I\left(\varkappa_{12}^{h}\right)\,dx_1\right]\,dx_1 + c_3(x_2) - c_2(x_2)\,x_1\,,\\
\tau_{22} &= \textstyle\int_0^{x_2}\left[f_2 - \int_0^{x_2} I\left(\varkappa_{12}^{h}\right)\,dx_2\right]\,dx_2 + c_4(x_1) - c_1(x_1)\,x_2\,,
\end{aligned} \tag{3.17}$$

where

$$\begin{aligned}
&c_0 \simeq \sigma_{12}\big|_{x_1=0,\,x_2=0}\,, && c_1(x_1) \simeq \partial\sigma_{12}/\partial x_1\big|_{x_2=0}\,, && c_2(x_2) \simeq \partial\sigma_{12}/\partial x_2\big|_{x_1=0}\,,\\
&c_3(x_2) \simeq \sigma_{11}\big|_{x_1=0}\,, && c_4(x_1) \simeq \sigma_{22}\big|_{x_2=0}\,,
\end{aligned} \tag{3.18}$$

and σ_{kl} may be understood as exact or FE values of the stresses.

There are many other ways of finding the appropriate stresses $\widetilde{\tau}_{kl}$ by means of the FE solution, which can be used as starting ones for evaluation of the equilibrated stress tensor $\boldsymbol{\tau}$. In particular, smoother interpolations may be more efficient.

In order to illustrate the essential difference from the approaches used by other authors, we formulate below basic a posteriori estimates in a form, in which the error in the smoothed FE stresses and the residual are separated.

Lemma 3.1. *Let* $\mathbf{v}$ *be arbitrary vector in* $\mathbb{V} = [\mathring{H}^1(\Omega)]^2$, $\boldsymbol{\sigma}(\mathbf{v})$ *be the stress tensor satisfying (3.2), (3.3) for* $\mathbf{u} = \mathbf{v}$ *and* $\mathbf{y} = \{y_{kl}\}_{k,l=1}^2$ *be an arbitrary symmetric tensor with the components in* $H^1(\Omega)$. *Then for* $\mathbf{u} - \mathbf{v}$ *either of the estimates*

$$\begin{aligned}
\|\mathbf{u} - \mathbf{u}_{\rm fem}\|_U &\le |\boldsymbol{\sigma}(\mathbf{u}_{\rm fem}) - \boldsymbol{\tau}|_\sigma\,,\\
\|\mathbf{u} - \mathbf{u}_{\rm fem}\|_U &\le |\boldsymbol{\sigma}(\mathbf{u}_{\rm fem}) - \mathbf{y}|_\sigma + |\delta\boldsymbol{\tau}|_\sigma\,,\\
\|\mathbf{u} - \mathbf{u}_{\rm fem}\|_U &\le |\boldsymbol{\sigma}(\mathbf{u}_{\rm fem}) - \mathbf{y}|_\sigma +\\
&\quad + \textstyle\sum_{k=1,2}\left\|\left(\frac{1-\nu^2}{E}\right)^{1/2}\int_{a_k(x_{3-k})}^{x_k}\left(f_k - \frac{\partial y_{k,k}}{\partial x_k} - \frac{\partial y_{1,2}}{\partial x_{3-k}}\right)(\eta_k, x_{3-k})\,d\eta_k\right\|_0\,,
\end{aligned} \tag{3.19}$$

holds, where $\boldsymbol{\tau}$ *is the stress tensor with the components*

$$\tau_{12} = y_{12}\,, \qquad \tau_{kk} = y_{kk}(a_k(x_{3-k})) + \int_{a_k(x_{3-k})}^{x_k}\left(f_k - \frac{\partial y_{1,2}}{\partial x_{3-k}}\right)(\eta_k, x_{3-k})\,d\eta_k$$

and

$$\delta\boldsymbol{\tau} = \begin{pmatrix} \delta\tau_{11} & 0\\ 0 & \delta\tau_{22}\end{pmatrix}\,, \qquad \delta\tau_{kk} = \int_{a_k(x_{3-k})}^{x_k}\left(f_k - \frac{\partial y_{k,k}}{\partial x_k} - \frac{\partial y_{1,2}}{\partial x_{3-k}}\right)(\eta_k, x_{3-k})\,d\eta_k\,.$$

For the stress tensor $\boldsymbol{\tau}$ one can also take one of the two tensors with the components, defined for $k = 1$ or $k = 2$ by formulas

$$\tau_{kk} = y_{kk}\,, \qquad \tau_{12} = y_{12}(a_k(x_{3-k})) + \int_{a_k(x_{3-k})}^{x_k} (f_k - \frac{\partial y_{k,k}}{\partial x_{3-k}})(\eta_k, x_{3-k})\, d\eta_k\,,$$

$$\tau_{3-k,3-k} = y_{3-k,3-k}(a_{3-k}(x_k)) + \int_{a_{3-k}(x_k)}^{x_{3-k}} (f_{3-k,3-k} - \frac{\partial y_{k,3-k}}{\partial x_k})(x_k, \eta_{3-k})\, d\eta_{3-k}\,.$$

Proof. Since all tensors $\boldsymbol{\tau}$, appearing in Lemma, satisfy the equilibrium equations, the first estimate (3.19) clearly holds. Tensor $\boldsymbol{\tau} = \mathbf{y} + \delta\boldsymbol{\tau}$ also satisfies the equilibrium equations. For this reason, the proof of the rest estimates requires only the inversion of the stress-strain relations and application of the Couchy and triangular inequalities. □

Assumptions of Lemma on the smoothness of tensor $\mathbf{y}$ can be easily sharpened and made matching the equilibrium equations in a weak sense (remind that in general tensor $\mathbf{y}$ itself does not satisfy them).

Remark 3.1. *Clearly, among three estimates (3.19), the most exact is the first one.*

3.2 Linear elasticity and more general problems of solid mechanics in 3-d

In the case of 3-d elasticity problem, the stress tensor

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}\,,$$

satisfying the equilibrium equations

$$\frac{\partial \sigma_{k1}}{\partial x_1} + \frac{\partial \sigma_{k2}}{\partial x_2} + \frac{\partial \sigma_{k3}}{\partial x_3} = f_k\,, \qquad k = 1, 2, 3\,. \tag{3.20}$$

can be obtained in a similar to the described above ways. For, instance we specify the shear stresses $\sigma_{12}, \sigma_{13}, \sigma_{23}$ by some sufficiently smooth functions, approximating the stresses, specified by the FE solution. Then the rest stresses are found from the equilibrium equations and their boundary values, specified either by the boundary conditions in stresses or by the approximate values, found by means of the FE solution. We will not describe these obvious algorithms and restrict ourselves to the formulation of a statement similar to Lemma 3.1.

For definiteness of the norms $|\cdot|_U$, $|\cdot|_\sigma$, $|\cdot|_\epsilon$ one can assume Hooke's law for the homogeneous linearly elastic body

$$\sigma_{kk} = \frac{E}{(1+\nu)(1-2\nu)}[(1-\nu)\varepsilon_{kk} + \nu(\varepsilon_{k+1,k+1} + \varepsilon_{k+2,k+2})]\,, \qquad \sigma_{kl} = \frac{E}{1+\nu}\varepsilon_{kl}\,, \quad k \neq l\,.$$

However, according to the above discussion, the first two estimates (3.21), given below, hold in a much more general situation under assumption of the proper definition of the norms $|\cdot|_U$, $|\cdot|_\sigma$.

Lemma 3.2. *Let* $\mathbf{v}$ *be arbitrary vector in* $\mathbb{V} = [\mathring{H}^1(\Omega)]^3$, $\boldsymbol{\sigma}(\mathbf{v})$ *be the stress tensor satisfying (3.20), (3.2) for* $\mathbf{u} = \mathbf{v}$ *and* $\mathbf{y} = \{y_{kl}\}_{k,l=1}^3$ *be an arbitrary tensor with the components in* $H^1(\Omega)$. *Then for* $\mathbf{u} - \mathbf{v}$ *either of the estimates*

$$
\begin{aligned}
|\mathbf{u} - \mathbf{u}_{\rm fem}|_U &\le |\boldsymbol{\sigma}(\mathbf{u}_{\rm fem}) - \boldsymbol{\tau}|_\sigma \,, \\
|\mathbf{u} - \mathbf{u}_{\rm fem}|_U &\le |\boldsymbol{\sigma}(\mathbf{u}_{\rm fem}) - \mathbf{y}|_\sigma + |\delta\boldsymbol{\tau}|_\sigma, \,, \\
|\mathbf{u} - \mathbf{u}_{\rm fem}|_U &\le |\boldsymbol{\sigma}(\mathbf{u}_{\rm fem}) - \mathbf{y}|_\sigma + \\
&\quad + \textstyle\sum_{k=1}^3 \| \frac{1}{\sqrt{E}} \int_{a_k}^{x_k} (f_k - \frac{\partial y_{k,k}}{\partial x_k} - \frac{\partial y_{k,k+1}}{\partial x_{k+1}} - \frac{\partial y_{k,k+2}}{\partial x_{k+2}})(\eta_k, x_{k+1}, x_{k+2})\, d\eta_k \|_0 \,,
\end{aligned}
\tag{3.21}
$$

holds, where $a_k = a_k(x_{k+1}, x_{k+2})$, $\boldsymbol{\tau}$ *is the stress tensor with the components*

$$
\tau_{kl} = y_{kl} \,, \qquad k \neq l \,,
$$

$$
\tau_{kk} = y_{kk}(a_k, x_{k+1}, x_{k+2}) + \int_{a_k}^{x_k} (f_k - \frac{\partial y_{k,k+1}}{\partial x_{k+1}} - \frac{\partial y_{k,k+2}}{\partial x_{k+2}})(\eta_k, x_{k+1}, x_{k+2})\, d\eta_k
$$

and

$$
\delta\boldsymbol{\tau} = \begin{pmatrix} \delta\tau_{11} & 0 & 0 \\ 0 & \delta\tau_{22} & 0 \\ 0 & 0 & \delta\tau_{33} \end{pmatrix} \,, \qquad \delta\tau_{kk} = \int_{a_k}^{x_k} (f_k - \frac{\partial y_{k,k}}{\partial x_k} - \frac{\partial y_{k,k+1}}{\partial x_{k+1}} - \frac{\partial y_{k,k+2}}{\partial x_{k+2}})(\eta_k, x_{k+1}, x_{k+2})\, d\eta_k \,.
$$

Proof. The proof is similar to the proof of Lemma 3.1. □

There are several other sequences of constructing symmetric equilibrated stress tensors. We can start from setting $\tau_{kl} = y_{kl}$ with arbitrary functions $y_{kl} \in H^1(\Omega)$ for three components τ_{kl}. Besides (1,2),(1,3),(2,3), there are other admissible combinations of k, l: (11),(22),(12); (22),(33),(23);(11),(33),(31);(12),(13),(22); (31),(32),(22). The rest stresses τ_{mp} are found from the equilibrium equations by 1-d integration. Boundary values, entering these integrals, are specified by arbitrary sufficiently smooth functions

$$
\tau_{mp}\big|_{\Gamma_{p,-}} = \tau_{mp}(a_p(x_{p+1}, x_{p+2}), x_{p+1}, x_{p+2}) \in H^{1/2}(\Gamma_{p,-}) \,.
$$

Let us underline that the equilibrium conditions do not depend on the type of the Hooke's law, *e.g.,* for orthotropic, transversally isotropic or other types of elastic bodies. As well they are not changed for a wide range of physically and geometrically nonlinear solid bodies. Therefore, the ways of obtaining of equilibrated and self-equilibrated stress tensors, introduced in this paper, are applicable to a wide range of problems in solid mechanics. All mentioned factors influence only techniques of evaluation of the (smoothed, if necessary) stress tensor, corresponding to the approximate solution, being subjected to a posteriori error estimation, and the specific energy norms, in which error estimation is produced.

Let us turn to a general case of nonlinear problems of solid mechanics, for which the approximate solutions obtained by means of the Lagrange principle of virtual work and Castigliano

principle of complementary work provide an upper and a lower bounds for the true potential energy of the body. As it is well known, in this case the a posteriori estimate can be written in the form

$$\mathcal{L}(\mathbf{v}) - \mathcal{L}(\mathbf{u}) \leq \mathcal{L}(\mathbf{v}) - \mathcal{C}(\boldsymbol{\tau})\,, \tag{3.22}$$

and under some conditions

$$\beta\|\mathbf{u}-\mathbf{v}\|_{\mathbb{V}} \leq \mathcal{L}(\mathbf{v}) - \mathcal{C}(\boldsymbol{\tau})\,, \qquad 0<\beta=\text{const}\,, \tag{3.23}$$

where

$\mathcal{L}(\mathbf{v})$ is the functional of the complete potential energy of the body on displacements $\mathbf{v}$, satisfying all geometric conditions,

$\mathbf{u}$ is the exact solution of the problem minimizing the functional $\mathcal{L}$,

$\mathcal{C}(\boldsymbol{\tau})$ is the functional of the complementary work on the stress tensor $\boldsymbol{\tau}$, satisfying the equilibrium conditions, and

$\|\cdot\|_{\mathbb{V}}$ is a norm satisfying the inequality

$$\beta\|\mathbf{u}-\mathbf{v}\|_{\mathbb{V}} \leq \mathcal{L}(\mathbf{v}) - \mathcal{L}(\mathbf{u})\,. \tag{3.24}$$

The estimate (3.22) expresses basic properties of the Lagrange and Castigliano principles see, *e.g.,* de Vebeke [15], Arthurs [4] Mosolov/Myasnikov [34], Washizu [54], Berdichevskii [8], where references on the earlier publications can be found. The estimate (3.23) follows from (3.22) under the condition that (3.24) is fulfilled. In a mathematical setting the basic facts for validity of the bounds (3.22),(3.23) may be found, *e.g.*, in Ekeland/Temam [14] and Duvaut/Lions[13], Glowinski [16]. They are found in the duality theory of the variational calculus and the theory of monotone/coercive operators. The latter allows to formulate conditions on the smoothness of data and the type of nonlinearity under which (3.24) holds.

The primal problem to be solved may be formulated in the following way: find $\mathbf{u}\in\mathbb{U}$ such that

$$\mathcal{L}(\mathbf{u}) = \inf_{v\in\mathbb{U}} \mathcal{L}(\mathbf{v})\,, \qquad \mathbb{U}\in\mathbb{V}\,, \tag{3.25}$$

where $\mathcal{L}$ is a proper convex, lower simicontinuous functional, $\mathbb{V}$ is a reflexive Banah space with the norm $\|\cdot\|_{\mathbb{V}}$ and $\mathbb{U}$ is a closed convex subset of $\mathbb{V}$. The variable $\boldsymbol{\tau}$ and the functional $\mathcal{C}(\boldsymbol{\tau})$ of the complementary work are the dual variable and the dual functional with respect to the primal variational problem (3.25). At that $\boldsymbol{\tau}$ belongs to the set $\mathbf{Q}_f = \mathbf{Q}_0 + \boldsymbol{\tau}_f$ of tensors satisfying equilibrium equations, *e.g.,* (3.20), with $\mathbf{Q}_0$ being the space of the self-equilibrated tensors. The problem of finding the stresses by means of the Castigliano principle is the dual problem: find $\boldsymbol{\sigma}\in\mathbf{Q} = \mathbf{Q}_0 + \boldsymbol{\tau}_u$, such that

$$\mathcal{C}(\boldsymbol{\sigma}) = \sup_{\boldsymbol{\tau}\in\mathbf{Q}} \mathcal{C}(\boldsymbol{\tau})\,, \qquad \mathbb{U}\in\mathbb{V}\,. \tag{3.26}$$

Assume that $(-\mathcal{C}(\boldsymbol{\tau}_0))$ is also a proper convex, lower simicontinuous functional, which is coercive on reflexive Banah space $\mathbf{Q}_0$, then

$$\mathcal{C}(\boldsymbol{\tau}) \leq \mathcal{C}(\boldsymbol{\sigma}) = \mathcal{L}(\mathbf{u}) \leq \mathcal{L}(\mathbf{v})\,, \qquad \forall\,\mathbf{v}\in\mathbb{V}\,,\ \forall\,\boldsymbol{\tau}\in\mathbf{Q}_0\,,$$

see,*e.g.,* Ekeland/Temam [14].

Many authors consider the use of the a posteriori estimates (3.22),(3.23) computationally very costly for the two reasons. One is that it is allegedly impossible to find the equilibrated tensor $\boldsymbol{\tau}$ close to the exact stress tensor $\boldsymbol{\sigma}$ in a direct and sufficiently cheap way. Another reason is based on the conviction that numerical solution of the dual problem for finding $\boldsymbol{\tau}$ is much more difficult that the numerical solution of the primal problem. Such reasons are soundly pronounced in some contemporary publications. However, in the preceding sections we have demonstrated that at least in some cases the first reason in a delusion. In what follows we consider some additional ways (including practical algorithms optimal in the arithmetic operations count) for finding equilibrated stress tensors close to the solution. Apart from that we will show that in many cases it is with no doubt feasible to develop numerical techniques for solving dual problems, which are comparable with the most efficient numerical techniques for solving primal problems in respect of the computational cost.

3.3 Examples of algorithms for numerical testing

Below we illustrate the described approach by the two algorithms for obtaining a posteriori estimates in the case of the linear plain strain elasticity problem problem (3.1)–(3.3),(3.5) in the square $\Omega = (0,1) \times (0,1)$. We use notations $\mathbf{e}_1 = (1,0)$, $\mathbf{e}_2 = (0,1)$, whereas $\boldsymbol{\sigma}_{\rm fem}$ stands for the stress tensor, determined by the FE solution.

Algorithm 3.1.

`Step 1.` For each point $y^{(i)} = (hi_1 + h/2, hi_2 + h/2),\ i_k = 0, 1, \ldots, n-1,$ calculate

$$\sigma_{\rm fem,12}(y^{(i)}) = \frac{E}{2(1+\nu)} \left(\frac{\partial u_{\rm fem,1}}{\partial x_2}(y^{(i)}) + \frac{\partial u_{\rm fem,2}}{\partial x_1}(y^{(i)}) \right),$$

using the finite element solution $\mathbf{u}_{\rm fem}$.

`Step 2.` Calculate approximate values $\hat{\phi}^{(i)}$ of $\partial\sigma_{\rm fem,12}/\partial x_2$ at the middle points $\hat{y}^{(i)} = (hi_1 + h/2, hi_2),$ of the horizontal mesh intervals:

$$\hat{\phi}^{(i)} = \frac{\sigma_{\rm fem,12}(y^{(i)}) - \sigma_{\rm fem,12}(y^{(i-\mathbf{e}_2)})}{h}, \qquad i_1 = 0, 1, .., n-1,\ i_2 = 1, 2, .., n-1\,.$$

`Step 3.` Evaluate approximate values $\phi^{(i)}$ of $\partial\sigma_{\rm fem,12}/\partial x_2$ at the internal nodes $x^{(i)}$:

$$\phi^{(i)} = \frac{1}{2}\left(\hat{\phi}^{(i-\mathbf{e}_1)} + \hat{\phi}^{(i)} \right), \qquad i_k = 1, 2, .., n-1\,.$$

`Step 4.` Calculate values $\phi^{(0,i_2)}$, $\phi^{(n,i_2)}$ for $i_2 = 1, 2, .., n-1$ by means of linear extrapolation from the interior of Ω over the two nearest values $\phi^{(i)}$ on the mesh line $x \equiv i_2 h$. In the same way calculate the values $\phi^{(i_1,0)}$ and $\phi^{(i_1,n)}$ for $i_1 = 1, 2, .., n-1$.

`Step 5.` Determine values of $\phi^{(i)}$ for $i_1, i_2 = 0, n$ corresponding the vertices, for instance, as the mean value of the two linear extrapolations along the two edges.

`Step 6.` Determine the piece wise bilinear continuous interpolation $I(\phi_h) \in \mathcal{V}(\Omega)$ of the mesh function $\phi_h = (\phi^{(i)})_{i_1,i_2=0}^{n}$.

`Step 7.` Evaluate components of the stress tensor $\boldsymbol{\tau}$ satisfying the equilibrium equation (3.11).

Step 7.1. For $x_k \in (0,1)$, define as piece linear continuous functions $c_{12}(x_1) \simeq \sigma_{\mathrm{fem},12}(x_1,0)$, $c_{11}(x_2) \simeq \sigma_{\mathrm{fem},11}(0,x_2)$, and $c_{22}(x_1) \simeq \sigma_{\mathrm{fem},22}(x_1,0)$. For instance, c_{12} is uniquely determined by its nodal values

$$c_{12}(i_1 h) = \sigma_{\mathrm{fem},12}(i_1 h, 0)\,, \qquad \text{for } i_1 = 0, n\,,$$

$$c_{12}(i_1 h) = \sigma_{\mathrm{fem},12}(i_1 - 0, 0) + \sigma_{\mathrm{fem},12}(i_1 + 0, 0)\,, \qquad \text{for } i_1 = 1, 2, .., n-1\,,$$

and similar procedures are used for c_{kk}.

Step 7.2. Determine components

$$\tau_{12} = c_{12}(x_1) + \int_0^{x_2} I(\phi_h)\, dx_2\,, \qquad \tau_{21} = \tau_{12}\,, \tag{3.27}$$

$$\tau_{11} = c_{11}(x_2) - \int_0^{x_1} (f_1 + I(\phi_h))\, dx_1\,, \tag{3.28}$$

$$\tau_{22} = c_{22}(x_1) - \int_0^{x_2} \left(f_2 + \frac{\partial \tau_{21}}{\partial x_1} \right) dx_2 \tag{3.29}$$

of the equilibrated stress tensor $\mathbf{t}$.

`Step 8.` Calculate $\|\boldsymbol{\sigma}(\mathbf{u}_{\mathrm{fem}}) - \boldsymbol{\tau}\|_\sigma$ for the a posteriori estimate.

Let us denote the stress tensor obtained by Algorithm 3.1 by $\mathbf{t}^{(1)}$. If to change variables x_1, x_2 for x_2, x_1, with the use of the same algorithm we come to another equilibrated stress tensor $\mathbf{t}^{(2)}$. Clearly, the tensor $\mathbf{t} = \alpha_1 \mathbf{t}^{(1)} + \alpha_2 \mathbf{t}^{(2)}$ also belongs to Q_f and

$$\|\mathbf{u} - \mathbf{u}_{\mathrm{fem}}\|_U \le \|\boldsymbol{\sigma}(\mathbf{u}_{\mathrm{fem}}) - (\alpha_1 \mathbf{t}^{(1)} + \alpha_2 \mathbf{t}^{(2)})\|_\sigma\,, \qquad \forall\ \alpha_1 + \alpha_2 \equiv 1\,. \tag{3.30}$$

Another invariant to variables x_1, x_2 procedure for finding a tensor τ, satisfying the equilibrium equations (3.11), follows Algorithm B and is presented in the algorithm 3.2.

Algorithm 3.2.

`Step 1.` Calculate the derivatives of the finite element solution $\mathbf{u}_{\mathrm{fem}} = (u_{\mathrm{fem},1}, u_{\mathrm{fem},2})$ at the centers of the mesh cells $y^{(i)} = (h(i_1 + 0.5), h(i_2 + 0.5))$, $i_k = 0, 1, \ldots, n-1$,

$$u_{\mathrm{fem},1,2}^{(i_1+1/2, i_2+1/2)} = \frac{\partial u_{\mathrm{fem},1}}{\partial x_2}(x_{i_1,i_2})\,, \quad u_{\mathrm{fem},2,1}^{(i_1+1/2, i_2+1/2)} = \frac{\partial u_{\mathrm{fem},2}}{\partial x_1}(x_{i_1,i_2}) \tag{3.31}$$

for $i_1, i_2 = 0, 1, \ldots, n-1$.

`Step 2.` Define values

$$\begin{aligned} \tilde{u}_{1,22}^{(i_1+1/2, i_2)} &= \frac{u_{\mathrm{fem},1,2}^{(i_1+1/2, i_2+1/2)} - u_{\mathrm{fem},1,2}^{(i_1+1/2, i_2-1/2)}}{h}\,, \\ \tilde{u}_{2,11}^{(i_1, i_2+1/2)} &= \frac{u_{\mathrm{fem},2,1}^{(i_1+1/2, i_2+1/2)} - u_{\mathrm{fem},2,1}^{(i_1-1/2, i_2+1/2)}}{h} \end{aligned} \tag{3.32}$$

for $i_1, i_2 = 0, 1, \ldots, n-1$.

`Step 3.` Define values

$$\tilde{u}^{(i_1,i_2)}_{1,221} = \frac{u^{(i_1+1/2,i_2)}_{1,22} - u^{(i_1-1/2,i_2)}_{1,22}}{h}\,, \quad \tilde{u}^{(i_1,i_2)}_{2,112} = \frac{u^{(i_1,i_2+1/2)}_{2,11} - u^{(i_1,i_2-1/2)}_{2,11}}{h} \tag{3.33}$$

for $i_1, i_2 = 1, 2, \ldots, n-1$.

`Step 4.` At the mesh nodes $(hi_1, 0)$ of the boundary for $i_1 = 1, 2, \ldots, n-1$, calculate $\tilde{u}^{(i_1,0)}_{1,221}$ with the use of the linear extrapolation by the two nearest values $\tilde{u}^{(i_1,1)}_{1,221}$ and $\tilde{u}^{(i_1,2)}_{1,221}$. Calculate

$$\tilde{u}^{(i_1,n)}_{1,221}\,, \quad \tilde{u}^{(0,i_2)}_{1,221}\,, \quad \tilde{u}^{(n,i_2)}_{1,221}\,, \quad \tilde{u}^{(i_1,0)}_{2,112}\,, \quad \tilde{u}^{(i_1,n)}_{2,112}\,, \quad \tilde{u}^{(0,i_2)}_{2,112}\,, \quad \tilde{u}^{(n,i_2)}_{2,112}$$

similarly. For the corner point $(0,0)$, determine $\tilde{u}^{(0,0)}_{1,221}$, *e.g.,* as the arithmetic mean of the two values obtained by linear extrapolations along two axes x_1, x_2 with the use of the two nearest values. Determine

$$\tilde{u}^{(0,n)}_{1,221}, \quad \tilde{u}^{(n,0)}_{1,221}, \quad \tilde{u}^{(n,n)}_{1,221}, \quad \tilde{u}^{(0,0)}_{2,112}, \quad \tilde{u}^{(0,n)}_{2,112}, \quad \tilde{u}^{(n,0)}_{2,112}, \quad \tilde{u}^{(n,n)}_{2,112}$$

similarly.

`Step 5.` Calculate

$$\tilde{\sigma}^{(i)}_{12,21} = \frac{E}{2(1+\nu)}\left(u^{(i)}_{1,221} + u^{(i)}_{2,112}\right)\,.$$

`Step 6.` Evaluate components of the stress tensor τ satisfying the equilibrium equation (3.11).

`Step 6.1.` Determine

$$\tau_{12} = \int_0^{x_2}\int_0^{x_1} I(\tilde{\sigma}_{12,21})\,dx_1dx_2 + c_0 + \int_0^{x_1} c_1(x_1)\,dx_1 + \int_0^{x_2} c_2(x_2)\,dx_2\,,$$

$\tau_{21} = \tau_{12},$ where $I(\tilde{\sigma}_{12,21})$ is the bilinear interpolation for $\tilde{\sigma}^{(i)}_{12,21}$ on the finite element mesh and

$$c_0 \simeq \sigma_{\mathrm{fem},12}(0,0)\,, \qquad c_1(x_1) \simeq \frac{\partial\sigma_{\mathrm{fem},12}}{\partial x_1}(x_1,0)\,, \qquad c_2(x_2) \simeq \frac{\partial\sigma_{\mathrm{fem},12}}{\partial x_2}(0,x_2)\,.$$

`Step 6.2.` For $c_3(x_2) \simeq \sigma_{\mathrm{fem},11}(0,x_2)$, determine

$$\tau_{11} = \int_0^{x_1}\left(f_1 - \int_0^{x_1} I(\tilde{\sigma}_{12,21})\,dx_1\right)dx_1 + c_3(x_2) - x_1c_2(x_2)\,. \tag{3.34}$$

`Step 6.3.` Define

$$\tau_{22} = \int_0^{x_2}\left(f_2 - \int_0^{x_2} I(\tilde{\sigma}_{12,21})\,dx_2\right)dx_2 + c_4(x_1) - x_2c_1(x_1),$$

where $c_4(x_1) \simeq \sigma_{\mathrm{fem},22}(x_1,0)$.

`Step 7.` Calculate the a posteriori estimator $|\boldsymbol{\sigma}(\mathbf{u}_{\mathrm{fem}}) - \boldsymbol{\tau}|_\sigma$.

We do not give formulas for evaluation of functions c_k, implying, however, averaging and interpolation procedures similar to those used in algorithms for Poisson equation. They provide the accuracy $\mathcal{O}(h^2)$ for stresses, if understood as direct approximations of stresses corresponding to smooth displacements. Let us emphasize that less accurate approximation of the boundary stresses, than inside of the domain, can damage the accuracy of the a posteriori estimator. The choice of c_k can be optimized on purpose to minimize the posteriori estimator. The system of algebraic equations for finding such c_k has by the order of h smaller dimension. This allows to arrange computations in such a way that the optimization will not compromise the optimality of the a posteriori estimator in the computational cost.[3]

4 Equilibrated fluxes/stresses obtained by means of Castigliano principle

To some authors, dual formulations of the boundary value problems, expressing in mechanics of solid bodies the Castigliano principle, seem difficult for numerical solution. By this reason dual formulations are often discarded from consideration as a tool in the process of the a posteriori estimation. However, we will illustrate that numerical solution of dual problems may be as simple as of primal problems.

4.1 Poisson equation

Solution of the problem (2.15) minimizes the functional

$$J(v) = \frac{1}{2}a(v,v) - (f,v)_\Omega\,, \qquad \forall\ v \in \mathbb{V} = \mathring{H}^1(\Omega)\,.$$

If to define Q_f as the set of functions satisfying the equilibrium equations in the generalized sense

$$Q_f = \{\mathbf{t} \in \mathbf{L}_2(\Omega)\ :\ \int_\Omega (\mathbf{t}\cdot\nabla v - fv) = 0,\ \forall\ v \in \mathbb{V}\}\,, \tag{4.1}$$

then the dual formulation of the problem is: find such $\mathbf{z} \in Q_f$ that

$$J^*(\mathbf{z}) = \min_{\mathbf{t}\in Q_f} J^*(\mathbf{t})\,, \qquad J^*(\mathbf{t}) = \frac{1}{2}\int_\Omega \mathbf{t}\cdot\mathbf{t}\,dx\,. \tag{4.2}$$

The solution of the dual problem can be represented in the form $\mathbf{z} = \mathbf{z}_0 + \mathbf{t}_f$, where $\mathbf{t}_f$ any vector from Q_f and the vector $\mathbf{z}_0 \in Q_0$ satisfies the equation

$$\int_\Omega (\mathbf{z}_0 - \mathbf{t}_f)\mathbf{t}_0\,dx = 0\,, \qquad \forall\ \mathbf{t}_0 \in Q_0\,. \tag{4.3}$$

For deriving a discrete approximation of the dual problem, we use the same as before square mesh of size $h = 1/n,\ n > 1,$ and define a subset $Q_f^h \subset Q_f$, which is represented as

$$Q_f^h = \mathbf{t}_f + Q_0^h\,,$$

[3]See, V.S. Kostylev. *A posteriori estimates optimal in the computational cost* . Master thesis. Chair of Applied Mathematics. St. Petersburg State Polytechnical University, St. Petersburg, Russia, 2006 (In Russian).

where Q_0^h is a finite dimensional subspace of Q_0 and $\mathbf{t}_f$ is any fixed vector from Q_f. The approximate solution $\mathbf{z}^h \in Q_f^h$ satisfies the equation

$$\int_\Omega (\mathbf{z}_0^h - \mathbf{t}_f)\mathbf{t}_0^h \, dx = 0\,, \qquad \forall\ \mathbf{t}_0^h \in Q_0^h\,. \tag{4.4}$$

In previous sections, we considered a number of ways of evaluation of the vector $\mathbf{t}_f$. For the purpose of discretization of the dual problem the simplest one can be used, *e.g.*, as in (2.7) at $q \equiv 0$. At the same time it is worth emphasizing that a good choice of $\mathbf{t}_f$ can considerably improve the a posteriori estimate, see also Remark 4.1. Suppose, $Q_0^h = \mathrm{span}\,[\boldsymbol{\phi}^{(i)}(x) = (\phi_1^{(i)}(x), \phi_2^{(i)}(x))^\top]_{i\in I^h}$, where I^h is the appropriate set of indices i, then (4.4) is reduced to solving the system

$$\mathbf{C}\mathbf{w} = \mathbf{f}\,, \tag{4.5}$$

with the matrix $\mathbf{C}$ and the vector $\mathbf{f}$ defined as

$$\begin{gathered} \mathbf{C} = \{c_{i,j}\}_{i,j\in I^h}\,, \qquad \mathbf{w} = \{w^{(i)}\}_{i\in I^h}\,, \qquad \mathbf{f} = \{f^{(i)}\}_{i\in I^h}\,, \\ c_{i,j} = \textstyle\int_\Omega \boldsymbol{\phi}^{(i)} \cdot \boldsymbol{\phi}^{(j)}\, d\,x\,, \qquad f^{(i)} = \int_\Omega \boldsymbol{\phi}^{(i)} \cdot \mathbf{t}_f \, d\,x\,. \end{gathered} \tag{4.6}$$

Remark 4.1. *Instead of the vector $\mathbf{t}_f$, one can use its approximation $\mathbf{t}_f^h$. Then the bound for the error of approximation $\|\mathbf{t}_f - \mathbf{t}_f^h\| = \mathcal{O}(h^\gamma)\mathbb{N}(f)$ with some γ and some norm $\mathbb{N}(f)$ of the function f will appear in the right part of the a posteriori estimate. The approximation $\mathbf{t}_f^h$ may be chosen on purpose,* e.g., *to simplify integration (since $\mathbf{t}_f^h$ can be obtained by approximation of f).*

The self-equilibrated vectors $\boldsymbol{\phi}^{(i)}$, $i_k = 1, 2, .., n,$ should be chosen in a way which lead to the system (4.5) with good computational properties. This can be anticipated, if they have localized supports. Each $\boldsymbol{\phi}^{(i)}$ introduced below has for the support the set $\varkappa_i = \omega_i \cap \Omega$, where $\omega_i = \{x : h(i_k - 1) < x_k < h(i_k + 1),\ k = 1, 2\}$. Before doing this, we remind the notation $\tau_i = \{x : h(i_k - 1) < x_k < hi_k\}$ for the square nests of the mesh. First we define subsidiary local functions $q^{(i)}$ which at the definition of self-equilibrated vectors play the same role as q in (2.7). Namely, we set

$$q^{(i)}(x) = \begin{cases} 1\,, & x \in \tau_i \cup \tau_{i_1+1,i_2+1}\,, \\ -1\,, & x \in \tau_{i_1+1,i_2} \cup \tau_{i_1,i_2+1}\,, \\ 0\,, & \Omega \setminus \overline{\omega}_i\,. \end{cases} \tag{4.7}$$

For the master linearly independent vectors, denoted as $\boldsymbol{\mu}^{(i)} = (\mu_1^{(i)}, \mu_2^{(i)})^\top$, and the vector $\mathbf{t}_f$, we take

$$\begin{gathered} \mu_1^{(i)}(x) = \int\limits_0^{x_1} q^{(i)}(\eta, x_2)\ d\eta\,, \qquad \mu_2^{(i)}(x) = -\int\limits_0^{x_2} q^{(i)}(x_1, \eta)\ d\eta\,, \\ \mathbf{t}_f(x) = (0, t_{f,2})^\top\,, \qquad t_{f,2}(x) = \int\limits_0^{x_2} f(x_1, \eta)\ d\eta\,. \end{gathered} \tag{4.8}$$

Clearly, the supports of these vectors are the sets ω_i, and instead of (4.8) one can use

$$\mu_1^{(i)}(x) = \int\limits_{(i_1-1)h}^{x_1} q^{(i)}(\eta, x_2)\ d\eta\,, \qquad \mu_2^{(i)}(x) = -\int\limits_{(i_2-1)h}^{x_2} q^{(i)}(x_1, \eta)\ d\eta\,. \tag{4.9}$$

The vectors $\boldsymbol{\phi}^{(i)}$ are defined as restrictions to $\varkappa^{(i)} = \omega^{(i)} \cap \Omega$ of the vectors $\boldsymbol{\mu}^{(i)}$, determined in (4.9), whereas $I^h = \{i : 0 \le i_k \le n\}$.

Now we compare the system (4.5), generated with the use of the coordinate vectors $\boldsymbol{\phi}^{(i)}$, and the FE systems for the Poisson equation. Namely we consider two FE systems

$$\mathbf{K}_{\rm D}\mathbf{u}_{\rm D,fem} = \mathbf{f}_{\rm D,fem}\,, \qquad \mathbf{K}_{\rm N}\mathbf{u}_{\rm N,fem} = \mathbf{f}_{\rm N,fem}\,, \tag{4.10}$$

generated by the spaces $\mathring{\mathcal{V}}(\Omega)$ and $\mathcal{V}(\Omega)$ for the Poisson equation in the unite square with the homogeneous Dirichlet and Neumann boundary conditions on its boundary, respectively. Remind that $\mathcal{V}(\Omega)$ is the space of the continuous piece wise bilinear functions and $\mathring{\mathcal{V}}(\Omega)$ is its subspace of functions vanishing on $\partial\Omega$. Let $\hat{p}^{(i)}(x)$ be the standard piece wise bilinear continuous function, satisfying conditions $\hat{p}^{(i)}(x^{(j)}) = \delta_{i,j}$ with $\delta_{i,j}$ being the Kronecker's delta, and $p^{(i)}(x)$ be its restriction to Ω. It is easy to conclude that

$$q^{(i)} = \frac{\partial^2 p^{(i)}}{\partial x_1 \partial x_2}\,, \qquad \boldsymbol{\phi}^{(i)} = (\frac{\partial p^{(i)}}{\partial x_2}, -\frac{\partial p^{(i)}}{\partial x_1})^\top\,, \tag{4.11}$$

whence it immediately follows that $\mathbf{C} = \mathbf{K}_{\rm N}$.

The system (4.10) with the matrix $\mathbf{K}_{\rm D}$ defines the FE solution of the primal problem. In order to obtain the equilibrated vector valid for the a posteriori estimation of the FE solution, one has to solve the second system (4.10) with the matrix $\mathbf{C} = \mathbf{K}_{\rm N}$. Clearly, the both can be solved very efficiently by many fast solvers, developed for FE methods.

Remark 4.2. *It is easy to note that the introduced localized fluxes satisfy the equality*

$$\sum_{i \in I^h} \boldsymbol{\phi}^{(i)} = 0\,, \qquad \forall\ x \in \Omega\,, \tag{4.12}$$

which is in agreement with to the fact that the matrix $\mathbf{C} = \mathbf{K}_{\rm N}$ *has the eigenvector* $\mathbf{y}_0 = \mathbf{1} = \{y_0^{(i)} = 1\}_{i \in I^h}$ *with the unity for all entries. This eigenvector corresponds to zero eigenvalue. According to (4.12), the right part in (4.5) satisfies the solvability condition, and if* $\mathbf{w}$ *is the solution, then* $\mathbf{w} + c\mathbf{1}$ *is also solution, where* c *is an arbitrary number.*

Suppose $\mathbf{K}$ is the FE matrix induced by (2.4), *i.e.,* it is the FE matrix for the problem (2.1),(2.2) with the mixed boundary condition. In this case, the matrix $\mathbf{C}$ coincides with the FE matrix for the problem

$$\begin{gathered} -\Delta u = f(x)\,, \qquad x \in \Omega\,, \qquad \partial\Omega = \overline{\Gamma}_D \cup \overline{\Gamma}_N\,, \\ u|_{\Gamma_N} = 0\,, \quad \partial u/\partial\nu|_{\Gamma_D} = 0\,, \end{gathered} \tag{4.13}$$

with Γ_D, Γ_N defined as in (2.2).

Remark 4.3. *The analogy between matrices* $\mathbf{C}$ *for the discretized dual problem and FE matrices* $\mathbf{K}$ *for the primal problem is retained for discretizations on rectangular grids by higher order rectangular finite elements. The situation is more complicated for triangulations by triangular elements compatible in C, because the second mixed derivative do not exist. However, if any*

elements of the class C^1 are used, then the analogy is retained. If the differential operator has the form $Lu = \sum_{k,l=1}^{2} \partial(a_{kl}\partial u/\partial x_k)/\partial x_l$. Coefficients in the dual formulation will be entries of the matrix $\mathbf{A}^{-1}$, where $\mathbf{A} = \{a_{kl}\}_{k,l=1}^{2}$ enters the primal formulation. In this case, the matrix $\mathbf{C}$ again will be the FE matrix, induced by the same coordinate functions as in the FE method for the primal formulation. However, $\mathbf{C}$ is generated for the elliptic problem with the coefficients defined by the matrix $\mathbf{A}^{-1}$.

Below we obtain another characterization of the of the space Q_0^h and the set of vectors $\{\boldsymbol{\phi}^{(i)}\}_{i\in I^h}$, used for generating system (4.5) with the matrix $\mathbf{C} = \mathbf{K}_{\rm N}$. This characterization will illuminate what kind of approximation is used for the dual problem. Yet another characterization by means of the integral equation with respect to the unknown function q will be considered in the next subsection.

We return again to the Dirichlet boundary value problem (2.26) in the unite square. Let

$$g^{(i)} = \begin{cases} 1 & \text{for } x \in \tau_i \cap \Omega\,, \\ 0 & \text{for } x \in \Omega \setminus \tau_i\,, \end{cases} \tag{4.14}$$

and

$$\widetilde{\mu}_{k0}^{i_{3-k}}(x_{3-k}) = \begin{cases} 1 & \text{for } x_{3-k} \in h(i_{3-k}-1, i_{3-k})\,, \qquad i_{3-k} = 1,2,..,n\,, \\ 0 & \text{for } x_{3-k} \notin h(i_{3-k}-1, i_{3-k})\,. \end{cases} \tag{4.15}$$

The set $\{g^{(i)}\}_{i_1,i_2=1}^{n}$ is the basis in the space $L_2^h(\Omega)$, which is a discrete approximation for $L_2(\Omega)$ and contains piece wise constant functions on the uniform square mesh of size h. At the definition of self-equilibrated vectors, the functions

$$g = \sum_{i_1,i_2=1}^{n} a^{(i)} g^{(i)}(x)$$

play the same role as q in (2.7). In turn, functions $\{\widetilde{\mu}_{k0}^{i_{3-k}}(x_{3-k})\}_{i_{3-k}=1}^{n}$ will serve as the basis for the approximation of the value of the boundary flux at $x_k \equiv 0$. In accordance with such understanding, we define the finite dimensional space of the self-equilibrated fluxes as $Q_0^h = \operatorname{span}[\mathbf{t}^{(i)}]_{i\in I_\star^h}$ where $I_\star^h = \{i = (i_k, i_{3-k}),\ i_k = 0,1,..,n,\ i_{3-k} = 1,2,..,n\}$ and

$$t_1^{(i)}(x) = \int_0^{x_1} g^{(i)}(\xi_1, x_2)\, d\xi_1\,, \qquad t_2^{(i)}(x) = -\int_0^{x_2} g^{(i)}(x_1, \xi_2)\, d\xi_2 \tag{4.16}$$

$$t_k^{(0,i_{3-k})}(x) = \widetilde{\mu}_{k0}^{i_{3-k}}(x_{3-k})\,, \qquad t_{3-k}^{(0,i_{3-k})}(x) \equiv 0\,. \tag{4.17}$$

It is easy to see that functions $q^{(i)},\ 1 \le i_k \le n$, are linear combinations of $g^{(i)}$ for the same i. Therefore, for these i fluxes $\boldsymbol{\phi}^{(i)}$ are linear combinations of fluxes $\mathbf{t}^{(i)}$. Fluxes $\boldsymbol{\phi}^{(0,i_{3-k})}$ are obtained by means of $q^{(i)},\ i_k = 0,\ 1 \le i_{3-k} \le n$ and boundary fluxes $\widetilde{\mu}_{k0}^{i_{3-k}}(x_k)$. Therefore, we have proved the following lemma.

Lemma 4.1. *The space $Q_0^h(\Omega) = \operatorname{span}[\mathbf{t}^{(i)}]_{i\in I_\star^h}$, spanned over fluxes (4.16),(4.17), and the space $Q_0^h(\Omega) = \operatorname{span}[\boldsymbol{\phi}^{(i)}(x)]_{i\in I^h}$, spanned over localized fluxes (4.9), coincide.*

Remark 4.4. *Solution of the system generated with the use of the coordinate functions* $\{\mathbf{t}^{(i)}\}_{i\in I_*^h}$ *may be unstable. The reason is not that this system has a bad condition number, which indeed is* $\mathcal{O}(h^{-2})$*. The system is equivalent to the discretized integral equation of the first kind – see next subsection – and at* $h \to 0$ *the smallest nozero eigenvalue of its matrix tends to zero.*

Suppose that we have to solve the problem (2.1),(2.2) for Poisson equation with the Neumann boundary condition on $\Gamma_N \neq \varnothing$. The approximate solution of the dual problem is represented as

$$\mathbf{z}^h = \mathbf{z}_0^h + \mathbf{t}_{f,N}\,, \qquad \mathbf{z}_0^h = \sum_{i_1,i_2=1}^{n-1,n} w^{(i)}\boldsymbol{\phi}^{(i)}\,,$$

where vector $\mathbf{t}_{f,N} \in Q_f$ should satisfy the balance equation and the Neumann boundary conditions whereas $\mathbf{z}_0^h$ belongs to the space of vectors satisfying homogeneous balance equation and homogeneous Neumann boundary conditions. In the case under consideration, the vector $\mathbf{t}_{f,N} \in Q_f$ can be defined as in the preceding case of the Dirichlet boundary condition, see (4.8). Coefficients, $w^{(i)}$ for $i \in I_0 := \{i : 1 \le i_1 \le n-1,\ 1 \le i_2 \le n\}$ are found from the system

$$\mathbf{C}\mathbf{w} = \mathbf{f}\,, \tag{4.18}$$

with the matrix $\mathbf{C}$ and the vectors $\mathbf{f}$ defined as

$$\begin{gathered}
\mathbf{C} = \{c_{i,j}\}_{i,j\in I_0}\,, \qquad \mathbf{w} = \{w^{(i)}\}_{i\in I_0}\,, \qquad \mathbf{f} = \{f^{(i)}\}_{i\in I_0}\,,\\
c_{i,j} = \textstyle\int_\Omega \boldsymbol{\phi}^{(i)}\cdot\boldsymbol{\phi}^{(j)}\,d\,x\,, \qquad f^{(i)} = \textstyle\int_\Omega \boldsymbol{\phi}^{(i)}\cdot\mathbf{t}_f\,d\,x\,,\\
\mathbf{t}_f(x) = (0, t_{f,2})^\top\,, \qquad t_{f,2}(x) = \int\limits_0^{x_2} f(x_1,\eta)\;d\eta\;.
\end{gathered} \tag{4.19}$$

In the particular case of the nonhomogeneous Neumann boundary condition (2.20),(2.21) the definition of $\mathbf{t}_f = (t_{f,1}, t_{f,2})$ should be changed. In particular, one can define $t_{f,1}$ as an arbitrary sufficiently smooth function satisfying boundary conditions on $\Gamma_{N,k},\ k = 1, 3$, and set

$$t_{f,2}(x) = g_2(x_1) + \int\limits_0^{x_2}\left(f - \frac{\partial t_{f,1}}{\partial x_1}\right)(x_1,\eta)\;d\eta\;.$$

Remark 4.5. *Suppose that we have to solve the problem (2.1) for the Poisson equation in an arbitrary domain* $\partial\Omega$ *with the Neumann boundary condition,* i.e., $\Gamma_N = \partial\Omega$ *and*

$$\partial u/\partial\nu\big|_{\partial\Omega} = g\,, \tag{4.20}$$

which for the dual formulation is essential. The solution $\mathbf{z}_{f,g}$ *of the dual problem can be represented in the form* $\mathbf{z} = \mathbf{z}_{0,0} + \mathbf{t}_{f,g}$*. Here* $\mathbf{t}_{f,g}$ *is any vector from the set* $Q_{f,g}$*, satisfying the balance equation and the boundary condition, and* $\mathbf{z}_{0,0}$ *is from the space* $Q_{0,0}$ *and satisfies the identity*

$$\int_\Omega (\mathbf{z}_{0,0} - \mathbf{t}_{f,g})\mathbf{t}_{0,0}\,dx = 0\,, \qquad \forall\ \mathbf{t}_{0,0} \in Q_{0,0}\,. \tag{4.21}$$

Since the main difficulty in defining $\mathbf{t}_{f,g}$ *is satisfying the boundary condition, for simplicity we assume* $f \equiv 0$. *One of the options is to use discrete approximation of* $\mathbf{t}_{0,g}$. *Let* Q_0^h *be the space defined as in the case of the Dirichlet boundary condition:* $Q_0^h = \operatorname{span}[\boldsymbol{\phi}^{(i)}]_{i\in I_\Omega^h}$ *with* $I_\Omega^h = \{i : \operatorname{mes}[\omega^{(i)} \cap \Omega \neq 0]\}$. *Note that among the set* $\{\boldsymbol{\phi}^{(i)}\}_{i\in I_\Omega^h}$ *not all elements are linearly independent, in what follows by* $\mathcal{I}_\Omega^h \subset I_\Omega^h$ *is understood such a subset that* $\{\boldsymbol{\phi}^{(i)}\}_{i\in\mathcal{I}_\Omega^h}$ *is the basis in* Q_0^h. *Let also* $Q_{\partial\Omega,0}^h = \operatorname{span}[\boldsymbol{\phi}^{(i)}]_{i\in I_{\partial\Omega}^h}$, *and*

$$\chi^{(i)}(x) = \phi_1^{(i)}\,\mathbf{e}_1\cdot\boldsymbol{\nu}(x) + \phi_2^{(i)}\,\mathbf{e}_2\cdot\boldsymbol{\nu}(x), \qquad \forall\, x\in\partial\Omega\,, \qquad \forall\, i\in I_{\partial\Omega}^h\,,$$

be the trace on $\partial\Omega$ *of the normal to the boundary component of the flux* $\boldsymbol{\phi}^{(i)}$. *We approximate* g *by*

$$g^h = \sum_{i\in I_{\partial\Omega}^h} b_i\boldsymbol{\chi}^{(i)}\,,$$

which by its definition is equilibrated. The approximation t_ν^h *may be found in different ways,* e.g., *by the collocation or the least square methods.*

Now we define the vector

$$\mathbf{t}_{0,g^h} = \sum_{i\in I_{\partial\Omega}^h} b_i\boldsymbol{\phi}^{(i)}\,,$$

belonging to the subspace Q_0^h *and satisfying approximately the equivalent to (4.20) boundary condition for the flux through the boundary. We can consider the Laplace equation with* g *in the boundary condition (4.20) replaced by* g^h, *denoting its exact solution by* u_{g^h}. *It can be solved also approximately on the basis of the Castigliano principle with the solution* $\mathbf{z}_{0,g^h} = \mathbf{z}_{0,0}^h - \mathbf{t}_{0,g^h}$ *satisfying*

$$\int_\Omega (\mathbf{z}_{0,0}^h - \mathbf{t}_{0,g^h})\mathbf{t}_{0,0}^h\,dx = 0\,, \qquad \mathbf{z}_{0,0}^h, \forall\ \mathbf{t}_{0,0}^h \in Q_{0,0}^h\,. \tag{4.22}$$

If $u_{\rm fem}$ *is the FE solution of the Laplace equation with the boundary condition (4.20), then*

$$\|\nabla(u_{g^h} - u_{\rm fem})\| \le \|\nabla u_{\rm fem} - \mathbf{z}_{0,g^h}\|\,. \tag{4.23}$$

At the same time

$$\|\nabla(u - u_{g^h})\|^2 \le \|u - u_{g^h}\|_{0,\partial\Omega}\,\|g - g^h\|_{0,\partial\Omega}$$

and, therefore, by the trace theorem and Friedrichs inequality

$$\|\nabla(u - u_{g^h})\|^2 \le c_{\rm trace}(1 + c_{\rm F})\|\nabla(u - u_{g^h})\|\,\|g - g^h\|_{0,\partial\Omega}\,.$$

Combining the last bound with (4.23), one gets the a posteriory estimate

$$\|\nabla(u - u_{\rm fem})\| \le \|\nabla u_{\rm fem} - \mathbf{z}_{0,g^h}\| + c_{\rm trace}(1 + c_{\rm F})\|g - g^h\|_{0,\partial\Omega}\,. \tag{4.24}$$

In this a posteriory estimate, instead of $\mathbf{z}_{0,g^h}$ *one can use* $\mathbf{t}_{0,g^h}$ *as well.*

Estimate (4.24) allows to consider a direct way of its implementation without attracting the dual formulation. First of all it can be rewritten in the form

$$\|\nabla(u-u_{\rm fem})\|_{0,\Omega} \le \|\nabla u_{\rm fem}-\mathbf{z}_0^h\|_{0,\Omega} + c_{\rm trace}(1+c_{\rm F})\|g-g^h(\mathbf{z}_0^h)\|_{0,\partial\Omega}\,, \qquad \forall\ \mathbf{z}_0^h\in Q_0^h\,, \quad (4.25)$$

in which $g^h(\mathbf{z}_0)$ is the normal component of the flux $\mathbf{z}_0$ on $\partial\Omega$. At the choice of the basis $\{\boldsymbol{\phi}^{(i)}\}_{i\in\mathcal{I}_\Omega^h}$ in Q_0^h, the conditions of the minimimum of the functional $F(\mathbf{z}_0^h)=\|\nabla u_{\rm fem}-\mathbf{z}_0^h\|^2_{0,\Omega},\ \forall\,\mathbf{z}_0^h\in Q_0^h$, result in the system of algebraic equations with the matrix having the 9-point stencil, similar to the FE matrix for the primal problem. One can also obtain an admissible vector $\mathbf{z}_0^h\in Q_0^h$ in a way of minimization of the sum $\sum_{y^{(j)}}(\nabla u_{\rm fem}-\mathbf{z}_0^h)^2\big|_{y^{(j)}}$ over a set of properly chosen points $y^{(j)}$, or in another way of approximation of FE flux $\nabla u_{\rm fem}$ by the vector from Q_0^h.

4.2 Some remarks and generalizations

Several obvious, but having important consequences, remarks can be made.

$\boldsymbol{\alpha}$) Since the Dirichlet boundary condition is natural for the dual formulation, we can use orthogonal grid for obtaining equilibrated fluxes in the case of arbitrary sufficiently smooth domain Ω.

$\boldsymbol{\beta}$) For solving the dual problem it is not necessary to use the same mesh, which was used for FE discretization. More over, since equilibrated fluxes from Q_0, see (4.1), are not supposed to satisfy compatibility conditions, we can easily add any equilibrated coordinate vectors to the basis $\{\boldsymbol{\phi}^{(i)}\}_{i\in\mathcal{I}_\Omega^h}$ and enrich the space Q_0^h. For instance, additional coordinate vectors can be defined with the use of local finer mesh, arbitrarily oriented with respect to the mesh used for the definition of the basis $\{\boldsymbol{\phi}^{(i)}\}_{i\in\mathcal{I}_\Omega^h}$ We can add also coordinate vectors with specific properties admitting a better approximation of concentration of fluxes or their singularities.

`Arbitrary domain.` Since the Dirichlet boundary condition is natural for the dual formulation, we can use orthogonal grid for obtaining equilibrated fluxes in the case of arbitrary sufficiently smooth domain Ω.

Suppose, we would like to obtain the equilibrated fluxes for the Dirichlet boundary value problem (2.26). Formally, the formulation of the dual problem is not changed, and again we have to solve integral identity (4.3) with the use of Q_0 defined for a given Ω as in (4.1). Namely, for finding the equilibrated fluxes which approximate the exact ones, we can cover the domain by the uniform square mesh of size h. Let $I_\Omega^h=\{i:\mathrm{mes}\,[\,\omega^{(i)}\cap\Omega\ne 0]\}$ and $Q_0^h=\mathrm{span}_{i\in I_\Omega^h}[\boldsymbol{\phi}^{(i)}]$, where each $\boldsymbol{\phi}^{(i)}$ is defined as the restriction of $\boldsymbol{\mu}^{(i)}$ to Ω. It is necessary to underline that since $\mathrm{mes}\,\boldsymbol{\varkappa}^{(i)}$ can be small for some $i\in I_\Omega^h$, the condition of matrix $\mathbf{C}$ of the system (4.5) for the problem under consideration can be bad. However, due to the discussed at the end of the preceding subsection analogy with the FE systems, several simple remedies for improving the condition can be used. They can be borrowed from those designed for solving primal problems with the natural boundary conditions by FE methods on the regular grids. We refer in this relation to Korneev [26] and Oganesian/Ruhovets [38].

In the case of the Neumann boundary condition on the whole boundary $\partial\Omega$, as in (4.20), a general and relatively cheap way of obtaining a posteriory estimate of the error is to use (4.25) with $\mathbf{z}_0^h$ defined at the end of Remark 4.5. The alternative way, which requires solution of the

dual problem (4.22) seems more difficult, for the reason that normal components of $\mathbf{t}_{0,g^h}$ and $\mathbf{z}^h_{0,0}$ on $\partial\Omega$ must be equal to g^h and 0. However, it is easy to come to conclusion that it is equivalent to solution by FE method of the Dirichlet problem with the explicitly given first boundary condition.

Lemma 4.2. *Let for simplicity $\partial\Omega$ be a continuously differentiable domain, $g \in L_\infty(\partial\Omega)$ and*

$$\int_{\partial\Omega} g d\,\bar{s} = 0\,, \qquad \varphi = \int_{s_0}^{s} g\,d\bar{s}\,.$$

where $d\,\bar{s}$ is the element of the length of the curve $\partial\Omega$. Let also u and w be the functions satisfying, respectively,

$$-\Delta u = 0\,, \qquad x \in \Omega\,, \qquad \partial\nu\big|_{\partial\Omega} = g\,, \qquad \int_{\partial\Omega} g\,ds = 0\,, \tag{4.26}$$

$$-\Delta w = 0\,, \qquad x \in \Omega\,, \qquad w|_{\partial\Omega} = \varphi\,, \qquad \varphi = \int_{s_0}^{s} g\,d\bar{s}\,. \tag{4.27}$$

Then

$$\frac{\partial u}{\partial x_1} = \frac{\partial w}{\partial x_2}\,, \qquad \frac{\partial u}{\partial x_2} = -\frac{\partial w}{\partial x_1}\,.$$

Proof. Indeed, if v is the arbitrary sufficiently smooth function, then $\mathbf{t} = (\partial v/\partial x_2, -\partial v/\partial x_1)^\top$ satisfies the balance equation, *e.g.*, for the problem (4.26). Besides, by the definitions of $\mathbf{t}$ and φ, the boundary condition $\mathbf{t}\cdot\boldsymbol{\nu}\big|_{\partial\Omega} = g$ is equivalent to $\partial v/\partial s = g$ or $v = c + \varphi$ with any constant c. Let $Q_0 = \{\mathbf{t} = (\partial v/\partial x_2, -\partial v/\partial x_1)^\top : \forall\, v \in H^1(\Omega)\}$ and $Q_{0,g} = \{\mathbf{t} \in Q_0 : \mathbf{t}\cdot\boldsymbol{\nu}\big|_{\partial\Omega} = g\}$. For so defined Q_0, we have $Q_0 = \{\int_\Omega \mathbf{t}\cdot\nabla\psi\,dx,\ \forall\ \psi \in \mathring{H}^1(\Omega)\}$, and, therefore, one can find the flux $\mathbf{z}_g$ for the problem (4.26) from the dual formulation

$$F(\mathbf{z}_g) = \min_{\mathbf{t}_g \in Q_{0,g}} F(\mathbf{t}_g)\,, \qquad F(\mathbf{t}_g) := \int_\Omega \mathbf{t}_g\cdot\mathbf{t}_g\,dx\,.$$

But by the definition of fluxes $\mathbf{t}_g \in Q_{0,g}$ and function φ, one has

$$\min_{\mathbf{t}_g \in Q_{0,g}} F(\mathbf{t}_g) = \min_{v\in H^1(\Omega),\, v|_{\partial\Omega}=\varphi} \int_\Omega \nabla v\cdot\nabla v\,dx\,,$$

completing the proof. □

From Lemma 4.2, it follows that for obtaining an a posteriory estimate we can proceed in the following way. If $w_{\rm fem} \in H^1(\Omega)$ is the FE solution of (4.27), we calculate $\mathbf{z}^h_0 = (\partial w_{\rm fem}/\partial x_2, -\partial w_{\rm fem}/\partial x_1)^\top$ and substitute it in (4.25). Note, that solutions of (4.27) by Galerkin's method and by other methods can be used with the same purpose. If the approximate solution $w_{\rm fem}$ is obtained by means of a discretization with noncompatible elements, one can smooth it and then use the smoothed FE solution instead of $w_{\rm fem}$.

`Densening of the mesh.` For solving the dual problem, it is convenient to use the same mesh, which was used for FE discretization (*e.g.,* for evaluation of the norms entering a posteriori error estimate), but not necessary. For instance, since equilibrated fluxes from Q_0, see (4.1), are

not supposed to satisfy compatibility conditions, we can easily add any equilibrated coordinate vectors to the set of such vectors, spanning the space Q_0^h, see, *e.g.,* (4.16),(4.17), and enrich this space up to some space $Q_{0,\star}^h \supset Q_0^h$. For instance, additional coordinate vectors can be defined with the use of local finer mesh. We can add also coordinate vectors with specific properties admitting a better approximation of singularities in fluxes. A good source of functions, which can serve for generating the localized equilibrated functions are coordinate functions used in *meshless methods*, see Oden/Duarte/Zienkiewcz [40] and Strouboulis/Babuska/Copps [47] and reach collection of references there. It seems also that coordinate functions of meshless methods have special advantages when used as generating functions $p^{(i)}$ in (4.11) for the coordinate vectors of self-balanced fluxes spanning the spaces Q_0^h and $Q_{0,\star}^h$ for solving dual problems.

We consider only a simple example of densening the mesh. Suppose that in a strictly internal subdomain $\mathcal{D} \subset \Omega$ it is necessary to use more accurate approximation. Since the case of an arbitrary sufficiently smooth domain was discussed above, we restrict considerations to the case when Ω is the unit square. We can proceed in the following way. Let $\mathcal{D}^h$ is the least "mesh domain" covering $\mathcal{D}$, *i.e.*,

$$\overline{\mathcal{D}}^h = \cup_{i \in I_D^h} \overline{\tau}_i^h \,, \qquad I_D^h = \{i : \tau_i^h \cap \mathcal{D} \neq \varnothing\} \,,$$

τ_i^h is the nest of the mesh of size h. Each square nest $\tau_i^h,\ i \in I_D$, is subdivided in four squares, defining the mesh of size $\hbar = h/2$ on $\mathcal{D}^h$. Retaining old indices $i = (i_1, i_2)$ for old nodes, we add indices $i = (i_1 \pm 1/2, i_2 \pm 1/2)$ for new nodes. We use the notation $\tau_i^\hbar$ for all smaller and bigger nests, assuming that i is the index of the right upper vertex of $\tau_i^\hbar$, and introduce sets

$$I_D^\hbar = \{i : \tau_i^\hbar \cap \mathcal{D}_h \neq \varnothing\} \,, \qquad I_\Omega^h = \{i : \tau_i^h \cap \Omega \neq \varnothing\} \,,$$

$$I^h_{\Omega \setminus \mathcal{D}^h} = \{i : \tau_i^h \cap \Omega \neq \varnothing\} \,, \qquad I^{h,\hbar} = I^h_{\Omega \setminus \mathcal{D}^h} \cup I_D^\hbar \,.$$

It is convenient to use the common notation τ_i for $\tau_i^h,\ i \in I^h_{\Omega \setminus \mathcal{D}^h}$ and for $\tau_i^\hbar,\ i \in I_D^\hbar$. Now we can directly use (4.14),(4.15) and (4.16),(4.17) for defining the coordinate vectors of equilibrated fluxes, spanning the space of equilibrated fluxes, which we denote $Q_0^{h,\hbar}$. The dimension of $Q_0^{h,\hbar}$ is $\mathrm{card}[I^{h,\hbar}] + 2n$.

The above consideration shows that in essence the densening is simple. However, the basis vectors (4.16),(4.17) are not localized and, therefore, the matrix of the corresponding system will have considerable fill in. Apart from that, solving this system may be unstable for the reason pointed out in Remark 4.4. The instability can be removed by the transformation of the introduced coordinate self-equilibrated fluxes to the localized ones. Instead of one type, two types of the localized self-equilibrated fluxes are used. If to use the notation $\boldsymbol{\phi}_h^{(i)} = \boldsymbol{\phi}^{(i)}$ for the fluxes, introduced in (4.7),(4.8),(4.9) on the mesh of size h, the second type fluxes $\boldsymbol{\phi}_{h/2}^{(j)}$ are similarly defined on the mesh of size $h/2$.

`Interpretation as solution of integral equation.` We turn to the Dirichlet problem (2.26) in an arbitrary sufficiently smooth domain and consider the equivalent system of three

equations

$$\begin{aligned} &\tfrac{\partial^2 u^{(1)}}{\partial x_1^2} = \alpha_1 f - q\,, && u^{(1)}\big|_{\partial\Omega} = 0\,, \\ &\tfrac{\partial^2 u^{(2)}}{\partial x_2^2} = \alpha_2 f + q\,, && u^{(2)}\big|_{\partial\Omega} = 0\,, \\ &u^{(1)} = u^{(2)}\,, && \forall\, x \in \Omega\,. \end{aligned} \tag{4.28}$$

In the mechanical sense, this system describes the two systems of strings stretched along axes x_k with each string of one direction fastened to the strings of other direction at the cross points. Function q is the internal force, acting between the two systems of strings. For simplicity we assume again that each line $x_k \equiv \text{const}$ crosses Ω not more than in two points, $\Gamma_{k,-}$ and $\Gamma_{k,+}$ are the parts of the boundary containing the points of such pairs, having lesser and larger coordinates x_k, respectively. We write the equations defining the curves $\Gamma_{k,-}$ and $\Gamma_{k,+}$ as $x_k = a_k(x_{3-k})$, and $x_k = b_k(x_{3-k})$ for $\hat{a}_{3-k} < x_{3-k} < \hat{b}_{3-k},$. Let $\mathcal{G}_k(x_k, x_{3-k}, y_k)$ be the Grin's functions for the ordinary differential operators in (4.28), so that

$$u^{(k)}(x) = \int_{a_k(x_{3-k})}^{b_k(x_{3-k})} \mathcal{G}_k(x_k, x_{3-k}, y_k)(\alpha_k f - q)(y_k, x_{3-k})\, d\, y_k\,.$$

Satisfying the equality $u^{(1)} = u^{(2)}$, one comes to the integral equation

$$\begin{aligned} &\textstyle\sum_{k=1,2} \int_{a_k(x_{3-k})}^{b_k(x_{3-k})} \mathcal{G}_k(x_k, x_{3-k}, y_k) q(y_k, x_{3-k})\, d\, y_k = \\ = \;&\textstyle\sum_{k=1,2} \int_{a_k(x_{3-k})}^{b_k(x_{3-k})} \mathcal{G}_k(x_k, x_{3-k}, y_k) \alpha_k f(y_k, x_{3-k})\, d\, y_k\,. \end{aligned} \tag{4.29}$$

In Rozin [45], Korneev/Rozin [29, 30], the class of integral equations of a more general but similar to (4.29) type was termed *integral equations of the method of splitting*.

For discretization of the integral equation (4.29), one can use the space $G^h(\Omega)$ of the piece wise constant functions with the basis $\{g^{(i)}\}_{i \in I^h_\Omega}$ defined as in (4.14), where $I^h_\Omega = \{i : \tau^h_i \cap \Omega \neq \varnothing\}$. As we will show below, in this way we come to the system equivalent to (4.5) up to the choice of the basis functions and α_k. However, from (4.29) it becomes clear that this system is not good for the numerical solution. Since (4.29) is an integral equation of the 1-st kind, this basis will lead to the unstable system of algebraic equations, the matrix of which has considerable fill in. The use of another set of coordinate fluxes $\{q^{(i)}\}$, which produce local self-equilibrated fluxes, results (as in Subsection 4.1) in the system, which computational properties are the same as of the FE system for the Poisson equation with the Neumann boundary condition. In order to make more clear the interrelation between solving procedures of the integral equation and the dual problem (4.2), we note that the generalized formulation of (4.29) is: find $q \in L_2(\Omega)$ such that for any $\widetilde{q} \in L_2(\Omega)$ we have

$$\begin{aligned} &\textstyle\int_\Omega \widetilde{q}(x) \{\sum_{k=1,2} \int_{a_k(x_{3-k})}^{b_k(x_{3-k})} \mathcal{G}_k(x_k, x_{3-k}, y_k) q(y_k, x_{3-k})\, d\, y_k\}\, d\, x = \\ = \;&\textstyle\int_\Omega \widetilde{q}(x) \{\sum_{k=1,2} \int_{a_k(x_{3-k})}^{b_k(x_{3-k})} \mathcal{G}_k(x_k, x_{3-k}, y_k) \alpha_k f(y_k, x_{3-k})\, d\, y_k\}\, d\, x\,. \end{aligned} \tag{4.30}$$

Taking into account the equalities

$$\int_{a_k(x_{3-k})}^{b_k(x_{3-k})} \mathcal{G}_k(x_k, x_{3-k}, y_k) q(y_k, x_{3-k})\, d\, y_k = 0 \qquad \text{for } x_k = a_k(x_{3-k}), b_k(x_{3-k})\,,$$

and integrating by parts, we obtain

$$\int_{a_k(x_{3-k})}^{b_k(x_{3-k})} \widetilde{q}(x)\{\int_{a_k(x_{3-k})}^{b_k(x_{3-k})} \mathcal{G}_k(x_k, x_{3-k}, y_k) q(y_k, x_{3-k})\, d\, y_k\}\, d\, x_k = -\int_{a_k(x_{3-k})}^{b_k(x_{3-k})} \{\psi(x_{3-k}) +$$

$$+ \quad \int_{a_k(x_{3-k})}^{x_k} \widetilde{q}(\eta_k, x_{3-k})\, d\eta_k\}\{\frac{\partial}{\partial x_k} \int_{a_k(x_{3-k})}^{b_k(x_{3-k})} \mathcal{G}_k(x_k, x_{3-k}, y_k) q(y_k, x_{3-k})\, d\, y_k\}\, d\, x_k\,,$$

where $\psi(x_{3-k})$ is an arbitrary sufficiently smooth function. By the definition of the Green's function $\mathcal{G}_k$, we have the equality

$$\frac{\partial}{\partial x_k} \int_{a_k(x_{3-k})}^{b_k(x_{3-k})} \mathcal{G}_k(x_k, x_{3-k}, y_k) p(y_k, x_{3-k})\, d\, y_k = -\widetilde{t}_{0,k}(x_{3-k}) - \int_{a_k(x_{3-k})}^{x_k} p(\eta_k, x_{3-k})\, d\eta_k\,,$$

in which $-\widetilde{t}_{0,k}(x_{3-k}) = \widetilde{t}_{0,k}(p(x), a_k(x_{3-k}), x_{3-k})$ is the boundary value for the derivative in the left part and, therefore, is uniquely defined by the function p. Suppose that for discretizing the problem we use the basis $\{g^{(i)}\}_{i \in I_\Omega^h}$ and

$$\begin{aligned} &\mathbf{t}_\circ^{(i)}(x) = (t_{\circ,1}^{(i)}, t_{\circ,2}^{(i)})^\top\,, \qquad t_{\circ,k}^{(i)} = \tfrac{\partial}{\partial x_k} \textstyle\int_{a_k(x_{3-k})}^{b_k(x_{3-k})} \mathcal{G}_k(x_k, x_{3-k}, y_k) q^{(i)}(y_k, x_{3-k})\, d\, y_k\,, \\ &\mathbf{t}_f(x) = (t_{f,1}^{(i)}, t_{f,2}^{(i)})^\top\,, \qquad t_{f,k} = \tfrac{\partial}{\partial x_k} \textstyle\int_{a_k(x_{3-k})}^{b_k(x_{3-k})} \mathcal{G}_k(x_k, x_{3-k}, y_k) \alpha_k f(y_k, x_{3-k})\, d\, y_k\,, \end{aligned} \tag{4.31}$$

$Q_\circ^h = \operatorname{span}[\mathbf{t}_\circ^{(i)}]_{i \in I_\Omega^h}$. Then the discrete form of (4.30) can be reformulated: find $\mathbf{z}^h = \mathbf{z}_\circ^h + \mathbf{t}_f$, where $\mathbf{t}_f$ was defined above and the vector $\mathbf{z}_\circ^h \in Q_\circ^h$ satisfies the equation

$$\int_\Omega (\mathbf{z}_\circ - \mathbf{t}_f)\mathbf{t}_\circ^{(i)}\, d\, x = 0\,, \qquad \forall\ i \in I_\Omega^h\,. \tag{4.32}$$

Let us underline that the coordinate functions $g^{(i)}$ of the internal forces are not self-equilibrated, but the fluxes $\mathbf{t}_\circ^{(i)}$ are due to their definition by means of the green's functions.

Remark 4.6. *Note that it is not necessary to use Green's functions $\mathcal{G}_k$ for evaluating vectors $\mathbf{t}_\circ^{(i)}, \mathbf{t}_f$. If the points $x_k^{\circ,i} = x_k^{\circ,i}(x_{3-k})$, $x_k^f = x_k^{f,k}(x_{3-k})$ are such that*

$$\int_{a_k(x_{3-k})}^{b_k(x_{3-k})} (x_k - x_k^{\circ,i}) q^{(i)}(x)\, d\, x_k = 0\,, \qquad \int_{a_k(x_{3-k})}^{b_k(x_{3-k})} (x_k - x_k^f)\alpha_k f(x)\, d\, x_k = 0\,,$$

then

$$\begin{aligned} &t_{\circ,k}^{(i)}(x) = t_{\circ,k,-}^{(i)}(x_{3-k}) + (-1)^{(3-k)} \textstyle\int_{a_k(x_{3-k})}^{x_k} q^{(i)}(y_k, x_{3-k})\, d\, y_k\,, \\ &t_{f,k}(x) = t_{f,k,-}(x_{3-k}) + \textstyle\int_{a_k(x_{3-k})}^{x_k} \alpha_k f(y_k, x_{3-k})\, d\, y_k\,, \end{aligned} \tag{4.33}$$

where

$$t_{\circ,k,-}^{(i)}(x_{3-k}) = \frac{b_k - x_k^{\circ,i}}{b_k - a_k} \int_{a_k(x_{3-k})}^{b_k(x_{3-k})} q^{(i)}(x)\, d\, x_k\,, \qquad t_{f,k,-}(x_{3-k}) = \frac{b_k - x_k^{\circ,i}}{b_k - a_k} \int_{a_k(x_{3-k})}^{b_k(x_{3-k})} \alpha_k f(x_k)\, d\, x_k\,.$$

The formulation (4.32) differs from (4.4) not only by the choice of the basis, but also in the spaces of coordinate functions. Indeed, we have $Q^h_\circ \subset Q^h_0$ and instead of (4.8) the relationships (4.33) are used. That means that vectors $t^{(i)}_{\circ,k}$, $t_{f,k}$ satisfy additional conditions in comparison with $t^{(i)}_{0,k}$, $t_{f,k}$ entering (4.4). If one defines

$$v^{(i)}_k(x) := \int_{a_k(x_{3-k})}^{x_k} t^{(i)}_{\circ,k}(y_k, x_{3-k})\, d\, y_k\,, \qquad w_k(x) := \int_{a_k(x_{3-k})}^{x_k} t_{f,k}(y_k, x_{3-k})\, d\, y_k\,,$$

then

$$v^{(i)}_k, w_k\big|_{\partial\Omega} = 0\,.$$

according to (4.31),(4.33) and definitions of Green's functions.

Remark 4.7. *In this paper, we do not discuss convergence of the described Riesz-Galerkin methods for the minimization of the functionals of the complementary work or corresponding integral equations of the splitting method. The analysis of the convergence does not meet difficulties. For the techniques which can be applied and results we refer to Korneev/Rozin [29, 30] and Korneev [19, 20, 21, 25, 27]. In these works, simpler for the realization, but more complicated for the analysis discretizations were studied. For instance, in Korneev [20, 21] 1-d integrals in (4.29) were approximated by the trapezium quadratures on a rectangular grid and then the collocation method was applied for obtaining the system of algebraic equations. Namely, it was required that (4.29) holds for the quadrature nodes. Let us also note, that integral equations obtained by the method of splitting for the bending problem of thin plates and cylindrical shells were studied by Korneev/Rozin [29, 30] and Korneev [22] – [25]. Construction of localized equilibrated functions of internal forces for thin shells and shells of moderate thickness was completed in Korneev [27] under rather general assumptions on the configuration of the middle surface. Also in [27] the analysis of the convergence may be found for numerical methods, based on the use of these equilibrated functions for the minimization of the functional of the complementary work for shells with arbitrary sufficiently smooth middle surfaces.*

Remark 4.8. *For the boundary value problems of mechanics of solid bodies two level of splitting are distinguished:* **i***) splitting of the equilibrium equations and* **ii***) splitting of partial differential equations of a boundary value problem in respect to displacements. The latter means that we are able to split the equilibrium, stress-strain, strain-displacements relations and boundary conditions. When this is possible, one can obtain integral equations of the method of splitting, which were introduced in mechanics of solid bodies by Rozin [45, 46]. From his and other works, mentioned in Remark 4.7, it follows that in rather general case* **ii***) can be accomplished, if the Poisson ratio ν is zero. However, for an efficient use of the a posteriori error estimation algorithms of the specific class under consideration one needs only to split equilibrium equations, which is always possible. The latter is true for the possibility of obtaining discrete dual formulations based on the Castigliano principle, which are comparable in the computational cost of their solution with FE methods for primal formulations in respect to displacements.*

Remark 4.9. *Let $\widetilde{u}_{\rm fem}$ be the function from the FE space interpolating the exact solution u. We rewrite (2.8) in the form*

$$\|\nabla u - \nabla u_{\rm fem}\| \le \|\nabla(\widetilde{u}_{\rm fem} - u_{\rm fem})\| + \|\nabla\widetilde{u}_{\rm fem} - \mathbf{z}\|\,, \qquad \forall\ \mathbf{z} \in Q_f\,, \tag{4.34}$$

and assume that the FE space is the space of the continuous piece wise bilinear functions. We assume also that $\mathbf{z}$ *is obtained by the approximate solution of the dual problem. Note that all functions in the right part are from the finite dimensional spaces. For convenience, let us call by the* gauge *order the order of convergence of the norm in the left, given by the a priori estimate. Due to the superconvergence property, the first term in the right can be estimated with an additional with respect to the gauge order, see,* e.g., *Oganesian/Ruhovets [38], Korneev [28] and Whalbin [53], Babuska/Strouboulis [7]. At the approximation of q by the piece wise constant functions it is easy to prove the estimate of the second norm with the gauge order. Therefore, the accuracy of the a posteriori error estimate is the same in the order as of the a priori estimate.*

`Comparison of discrete primal and dual formulations.` The equality $\mathbf{C}_I = \mathbf{K}$ takes place in a much more general case. To illustrate this we turn to the Dirichlet problem (2.26) in an arbitrary sufficiently smooth domain and its FE discretization (2.27). We can assume that the finite elements of the FE assemblage are arbitrary which can provide that $\mathring{\mathcal{V}}(\Omega) \in C(\overline{\Omega}) \cap \mathring{H}^1(\Omega)$. In other words, the finite elements are allowed to be be curvilinear and associated with the triangular or rectangular reference element with any compatible in $C(\overline{\Omega}) \cap \mathring{H}^1(\Omega)$ shape functions. The Hermite finite elements are not excluded, but we number the FE Galerkin basis functions of the space $\mathcal{V}(\Omega)$ consecutively with the use of the number $l = 1, 2, .., \mathcal{L}$ without making difference between basis functions, corresponding to the values of FE functions or their derivatives at the nodes. Therefore, $\mathcal{L}$ is the total number of the FE Galerkin basis functions, for which we use now the notation $p^{[l]}(x)$. The number of the internal basis functions is denoted by $\mathcal{L}_I$ so that and $\mathring{\mathcal{V}}(\Omega) = \mathrm{span}[p^{[l]}]_{l=1}^{\mathcal{L}_I}$. The finite element solution satisfies the identity (2.27). The basis self-equilibrated vectors $\boldsymbol{\phi}_0^{[l]} = (\phi_{0,1}^{[l]}, \phi_{0,2}^{[l]})^\top$ in the space $Q_0 = \mathrm{span}[\boldsymbol{\phi}_0^{[l]}]_{l=1}^{\mathcal{L}}$ can be defined by means of the FE basis functions according to

$$\phi_{0,k}^{[l]} = (-1)^{1-k} \frac{\partial p^{[l]}}{\partial x_{3-k}}\,, \qquad k = 1, 2\,, \quad l = 1, 2, .., \mathcal{L}_I\,. \tag{4.35}$$

It is clear that vectors $\boldsymbol{\phi}_0^{[l]}$ satisfy the equilibrium equation

$$\frac{\partial \phi_{0,1}^{[l]}}{\partial x_1} + \frac{\partial \phi_{0,2}^{[l]}}{\partial x_2} = \frac{\partial^2 p^{[l]}}{\partial x_1 \partial x_2} - \frac{\partial^2 p^{[l]}}{\partial x_2 \partial x_1} = 0 \tag{4.36}$$

in classical sense only when the finite elements are compatible in C^1. In general, for elements compatible in C, second derivatives in (4.36) are Dirack's deltas on the borders of finite elements, and therefore (4.36) involves equalities for the Dirack's deltas corresponding to $\partial^2 p^{[l]} / \partial x_k \partial x_{3-k}$ for $k = 1, 2$. However, in the weak sense, see (4.1), the equilibrium equations are satisfied. Now we see that $\mathbf{K}_I = \mathbf{C}_I$, where

$$\begin{gathered} \mathbf{K} = \{k_{l,m}\}_{l,m \in \mathcal{L}_I}\,, \qquad \mathbf{C} = \{c_{l,m}\}_{l,m \in \mathcal{L}_I}\,,\,, \\ k_{l,m} = \int_\Omega \nabla p^{[l]} \cdot \nabla p^{[m]}\, dx\,, \qquad c_{l,m} = \int_\Omega \boldsymbol{\phi}_0^{[l]} \cdot \boldsymbol{\phi}_0^{[m]}\, d\,x\,. \end{gathered} \tag{4.37}$$

Clearly, in the case of more general equations of second order, *e.g.*, when the coefficients are different and variable, the matrices $\mathbf{K}_I$ and $\mathbf{C}_I$ are not equal. For instance, if we turn to the

equation $\nabla \cdot \boldsymbol{\rho} \nabla u = f$ where $\boldsymbol{\rho}$ is a diagonal 2×2 matrix $\boldsymbol{\rho} = \mathrm{diag}[\rho_1, \rho_2]$, then $\mathbf{K}_I$ is the same as $\mathbf{C}_I$ for the similar equation with $\boldsymbol{\rho} = \mathrm{diag}[\rho_2^{-1}, \rho_1^{-1}]$. Indeed, the basis in Q_f is the same, and the coefficients of the stiffness and deflection matrices are the integrals

$$k_{l,m} = \int_\Omega \nabla p^{[l]} \cdot \boldsymbol{\rho} \nabla p^{[m]} \, dx \,, \qquad c_{l,m} = \int_\Omega \boldsymbol{\phi}_0^{[l]} \cdot \boldsymbol{\rho}^{-1} \boldsymbol{\phi}_0^{[m]} \, d\, x \,.$$

From the above considerations, one can conclude that at least for regular elliptic problems computational properties of the discretizations of primal and dual problems are in essential the same and solutions of these diecretizations can be obtained by fast solvers of the same types.

Remark 4.10. *For 3d Dirichlet problem (2.15) we have for the flux* $\mathbf{t} = (t_1, t_2, t_3)^\top$ *the balance equation*

$$\frac{\partial t_1}{\partial x_1} + \frac{\partial t_2}{\partial x_2} + \frac{\partial t_3}{\partial x_3} + f = \mathbf{0} \,. \tag{4.38}$$

Therefore, in order to obtain an equilibrated flux, two of the components can be specified by arbitrary functions and only third found from (4.38). Again, sufficiently smooth local functions can be used for generating self-balanced fluxes. Suppose for simplicity that $\phi(x)$ *has local support* δ *and* $\partial^3 \phi / \partial x_1 \partial x_2 \partial x_3$ *is bounded in the vicinity of* δ*. Then for sufficiently smooth functions* α_k, $k = 1, 2, 3$, $\alpha_1 + \alpha_2 + \alpha_3 = 0$, *components of a self-balanced flux are defined by*

$$t_k = \alpha_k \frac{\partial^2 t_k}{\partial x_{k+1} \partial x_{k+2}} \,, \qquad k = 1, 2, 3 \,. \tag{4.39}$$

Suppose, $\mathcal{V}(\Omega)$, $\Omega = (0,1)^3$, *is the space of the continuous piece wise trilinear functions on the FE cubic mesh of size* h *and* $\phi^{(i)}(x)$ *satisfies* $\phi^{(i)}(x^{(j)}) = \delta_{i,j}$, *where* $i = (i_1, i_2, i_3), j = (j_1, j_2, j_3)$, $x^{(i)} = hi$, $x^{(j)} = hj$ *and* $h = 1/n$. *Then substituting* $\phi(x) = \phi^{(i)}(x)$ *in (4.39), one obtains local self-balanced fluxes* $\mathbf{t}^{(i)} = (t_1^{(i)}, t_2^{(i)}, t_3^{(i)})^\top$.

4.3 Linear elasticity problems

`State of plain stress.` In this section we will use the matrix-vector form of the stress strain relations for the state of plain stress

$$\boldsymbol{\sigma} = \mathbf{D} \boldsymbol{\varepsilon} \,, \qquad \mathbf{D} = \frac{E}{(1+\nu)(1-2\nu)} \begin{pmatrix} 1-\nu & \nu & 0 \\ \nu & 1-\nu & 0 \\ 0 & 0 & (1-2\nu)/2 \end{pmatrix} \,,$$

where

$$\boldsymbol{\sigma} = (\sigma_{11}, \sigma_{22}, \sigma_{12})^\top, \qquad \boldsymbol{\varepsilon} = (\varepsilon_{11}, \varepsilon_{22}, \gamma_{12})^\top \,, \qquad \gamma_{12} = 2\varepsilon_{12} \,,$$

and we turn to the problem (3.6). The solution of the dual problem is the stress vector $\mathbf{z} = \mathbf{z}_0 + \mathbf{t}_f$, where $\mathbf{t}_f$ any vector from $\mathbf{Q}_f$ and the vector $\mathbf{z}_0 \in \mathbf{Q}_0$ satisfies the equation

$$\int_\Omega (\mathbf{z}_0 - \mathbf{t}_f) \mathbf{D}^{-1} \mathbf{t}_0 \, dx = 0 \,, \qquad \forall \ \mathbf{t}_0 \in \mathbf{Q}_0 \,. \tag{4.40}$$

In order to discretize this integral identity, one can proceed along the lines of Algorithms A or B. Suppose for simplicity that the domain is covered by the square mesh of size h, Ω^h is some mesh domain containing Ω and defined below, and

$$\mathcal{V}(\Omega^h) = \{v : v \in C(\overline{\Omega}^h),\ v\big|_{\tau_i} \in \mathcal{Q}_p,\ p \geq 1\}\,,$$

where $\mathcal{Q}_p$ is the space of polynomials of the the order p in each variable. For defining a discrete subspace $\mathbf{Q}_0^h(\Omega) \subset \mathbf{Q}_0(\Omega)$, at first we define the space of stresses σ_{12} as the restriction $\mathcal{V}(\Omega)$ of the space $\mathcal{V}(\Omega^h)$ to Ω. The stresses σ_{kk} are evaluated according (3.13),(3.14) with $f_k \equiv 0$ and functions $\psi_{kk,\Gamma_{k,-}}$ from appropriate finite dimensional spaces of traces. The coefficients before the basis functions of these spaces are clearly additional unknowns in the discrete dual formulation. It is possible to avoid special description of functions $\psi_{kk,\Gamma_{k,-}}$ by choosing a proper basis in $\mathbf{Q}_0^h(\Omega)$. At the same time, it is possible to define the basis functions in such a way that their supports will be localized on the squares, containing 9 mesh sells. Since the generalization to any $p \geq 1$ is obvious, we will describe one of the bases for the case $p = 1$.

Let $p^{(i)}(x),\ p^{(i)}(x^{(j)}) = \delta_{i,j}$ be the usual nodal coordinate function in the space of continuous piece wise bilinear functions, and

$$\phi^{(i)}(x) = p^{(i)}(x) - p^{(i_1-1,i_2)}(x) - p^{(i_1,i_2-1)}(x) + p^{(i_1-1,i_2-1)}(x)\,, \tag{4.41}$$

so that $\mathrm{supp}\,[\phi^{(i)}(x)] = \omega^{(i)}$, and $\omega^{(i)} = \{x : (i_k - 1)h < x_k < (i_k + 1)h\}$. First, we define the master basis vector $\boldsymbol{\mu}^{(i)}$ by the equalities

$$\mu_{12}^{(i)} = \phi^{(i)}(x)\,, \qquad \mu_{kk}^{(i)} = -\int_{(i_k-1)h}^{x_k} \frac{\partial \phi^{(i)}}{\partial x_{3-k}}(\eta_k, x_{3-k})\, d\eta_k\,,$$

Elements of the basis $\{\boldsymbol{\phi}_0^{h,i}\}_{\mathcal{I}^h}$, $\mathcal{I}^h \subseteq I^h = \{i : \varkappa^{(i)} := \omega^{(i)} \cap \Omega \neq \varnothing\}$, in $\mathbf{Q}_0^h(\Omega)$ are the restrictions of $\boldsymbol{\mu}^{(i)}$ to Ω. Note that for Ω^h one can take the domain with the closure $\overline{\Omega}^h = \cup_{I^h} \overline{\omega}^{(i)}$. Discrete formulation of (4.40) is the following one: find such vector $\mathbf{z} = \mathbf{z}_0^h + \mathbf{t}_f$ with $\mathbf{z}_0^h \in \mathbf{Q}_0^h(\Omega)$ that

$$\int_\Omega (\mathbf{z}_0^h - \mathbf{t}_f)\mathbf{D}^{-1}\boldsymbol{\phi}_0^{h,i}\, dx = 0\,, \qquad \forall\ \boldsymbol{\phi}_0^{h,i} \in \mathbf{Q}_0^h\,. \tag{4.42}$$

`3-d elasticity.` As it is seen from Section 3, the construction of the space of the self-equilibrated stresses for 3-d is similar to 2-d case. The same is true about the construction of the basis functions in the space Q_0^h, which have local supports. In the matrix-vector form the stress strain relations are

$$\boldsymbol{\sigma} = \mathbf{D}\boldsymbol{\varepsilon}\,,$$

$$\mathbf{D} = \tfrac{E}{(1+\nu)(1-2\nu)} \begin{pmatrix} 1-\nu & \nu & \nu & 0 & 0 & 0 \\ & 1-\nu & \nu & 0 & 0 & 0 \\ & & 1-\nu & 0 & 0 & 0 \\ & & & (1-2\nu)/2 & 0 & 0 \\ & SYM & & & (1-2\nu)/2 & 0 \\ & & & & & (1-2\nu)/2 \end{pmatrix}, \tag{4.43}$$

where

$$\boldsymbol{\sigma} = (\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{12}, \sigma_{23}, \sigma_{13})^\top, \qquad \boldsymbol{\varepsilon} = (\varepsilon_{11}, \varepsilon_{22}, \varepsilon_{33}, \gamma_{12}, \gamma_{23}, \gamma_{12})^\top, \qquad \gamma_{kl} = 2\varepsilon_{kl}, \quad k \neq l.$$

As for the 2-d problem, we cover the domain by the square mesh of size h, consider some domain $\Omega^h \supseteq \Omega$ containing Ω, and the space $\mathcal{V}(\Omega^h) = \{v : v \in C(\overline{\Omega}^h),\ v\big|_{\tau_i} \in \mathcal{Q}_p,\ p \geq 1\}$, where $\mathcal{Q}_p$ is the space of polynomials of the the order p in each variable. For Ω^h we take the mesh domain containing all nests $\tau_i,\ i = (i_1, i_2, i_3)$ involved in the definition of the basis in the space $\mathbf{Q}_0^h$.

Let $p^{(i)}(x),\ p^{(i)}(x^{(j)}) = \delta_{i,j}$ be the usual nodal coordinate function in the space of continuous piece wise bilinear functions, and

$$\begin{aligned}\phi^{(i)}(x) \quad &= p^{(i)}(x) - p^{(i_1-1,i_2,i_3)}(x) - p^{(i_1,i_2-1,i_3)}(x) + p^{(i_1-1,i_2-1,i_3)}(x) - \\ &-p^{(i_1,i_2,i_3-1)}(x) + p^{(i_1-1,i_2,i_3-1)}(x) + p^{(i_1,i_2-1,i_3-1)}(x) - p^{(i_1-1,i_2-1,i_3-1)}(x),\end{aligned} \tag{4.44}$$

so that $\operatorname{supp}[\phi^{(i)}(x)] = \omega^{(i)},$ and $\omega^{(i)} = \{x : (i_k - 2)h < x_k < (i_k + 1)h,\ k = 1, 2, 3\}$.

We define the master basis vector $\boldsymbol{\mu}^{(i)}$ by the equalities

$$\mu_{kl}^{(i)} = \phi^{(i)}(x), \qquad \mu_{kk}^{(i)} = -\int_{(i_k-2)h}^{x_k} (\frac{\partial \phi^{(i)}}{\partial x_{k+1}} + \frac{\partial \phi^{(i)}}{\partial x_{k+2}})(\eta_k, x_{k+1}, x_{k+2})\, d\eta_k, \qquad k \neq l,$$

and the elements of the basis $\{\boldsymbol{\phi}_0^{h,i}\}_{I^h}$, $I^h = \{i : \varkappa^{(i)} := \omega^{(i)} \cap \Omega \neq \varnothing\}$, in $\mathbf{Q}_0^h(\Omega)$ as the restrictions of $\boldsymbol{\mu}^{(i)}$ to Ω.

The discrete formulation of the dual problem has the same form (4.42) and requires solution of the system of linear algebraic equations with the banded matrix.

5 Numerical results

In this section, we discuss the results of numerical experiments with the equilibrium based a posteriori error estimates described in previous sections. The purpose of our experiments is to demonstrate that our algorithms are able to produce such estimates with the very good effectiveness index and for the optimal in the order number of arithmetic operations. For model problems, linear and nonlinear second order elliptic equations in the unite square were used, including the equation with jumping coefficients and the plain strain linear elasticity problem. Main conclusions made from numerical results are that our a posteriori estimates

are *asymptotically exact*, and, more over, in many cases convergence of the effectiveness index to the unity was observed at $h \to 0$,

are *asymptotically optimal in the computational cost*, the number of arithmetic operations was always proportional to the number of unknowns,

can be easily made *robust* in respect to coefficients jumps after necessary modifications of the algorithms.

We tested also the a posteriori estimators in which the equilibrated fields were obtained by solving the dual problem, expressing the Castigliano principle.

5.1 The Poisson equation

5.1.1 Direct evaluation algorithms

Consider the model problem

$$
\begin{aligned}
&-\Delta u = \frac{5\pi^2}{2}\cos\frac{3\pi}{2}x_1\cos\frac{\pi}{2}x_2\ , \quad (x_1,x_2)\in\Omega=(0,1)\times(0,1)\ ,\\
&u|_{\Gamma_D}=0\ , \quad u|_{\Gamma_N}=0\ ,\\
&\Gamma_D=\{(x_1,x_2)|\ x_1\in[0,1]\ ,x_2=1\}\cup\{(x_1,x_2)|\ x_1=1\ ,x_2\in[0,1]\}\ ,\\
&\Gamma_N=\{(x_1,x_2)|\ x_1\in[0,1]\ ,x_2=0\}\cup\{(x_1,x_2)|\ x_1=0\ ,x_2\in[0,1]\}\ .
\end{aligned}
\tag{5.1}
$$

with the exact solution

$$
u(x_1,x_2)=\cos\frac{3\pi}{2}x_1\cos\frac{\pi}{2}x_2\ . \tag{5.2}
$$

The FE space of the piece wise bilinear functions for this problem $\mathcal{V}_0(\Omega)$ was defined in Section 2. For the FE solution $u_{\rm fem}$ on the mesh of size h, we used Algorithm 2.1 in order to calculate the vector-valued function $\mathbf{t}(x)=(t_1(x),t_2(x))^\top$, which satisfies the balance equation (2.5) and the boundary conditions (2.6). Then we calculated the energy norm of the error $e=\|\nabla(u-u_{\rm fem})\|$ and the a posteriori estimator $\eta=\|\nabla u_{\rm fem}-\mathbf{t}\|$, *i.e.,* the left and right sides correspondingly in (2.8).

Figure 1 shows the dependence of the energy norm of the FE error and of the a posteriori estimator on the number of unknowns N. It demonstrates the same asymptotic behavior of the both values. The number of unknowns in this experiment exceeded $4\cdot10^6$, but these lines practically coincide for N greater than 10^4. The a posteriori estimator η stays greater than the energy norm of the error e (see Table 1). This validates that the equilibrium based a posteriori estimate guarantees the upper asymptotically exact estimate.

N	e	η	$I_{\rm eff}$
16	$8.40422\cdot10^{-1}$	$2.87729\cdot10^{-1}$	3.42362
64	$4.08785\cdot10^{-1}$	$8.38575\cdot10^{-1}$	2.05138
256	$2.02318\cdot10^{-1}$	$2.67228\cdot10^{-1}$	1.32083
1024	$1.00877\cdot10^{-1}$	$1.09368\cdot10^{-1}$	1.08417
4096	$5.04023\cdot10^{-2}$	$5.14654\cdot10^{-2}$	1.02109
16384	$2.51966\cdot10^{-2}$	$2.53287\cdot10^{-2}$	1.00525
65536	$1.25977\cdot10^{-2}$	$1.26141\cdot10^{-2}$	1.00131
262144	$6.29880\cdot10^{-3}$	$6.30085\cdot10^{-3}$	1.00033
1048576	$3.14939\cdot10^{-3}$	$3.14965\cdot10^{-3}$	1.00008
4194304	$1.57469\cdot10^{-3}$	$1.57473\cdot10^{-3}$	1.00002

Table 1.

Figure 2 shows the dependence of ($I_{\rm eff}-1$ on N), where

$$
I_{\rm eff}=\frac{\eta}{e} \tag{5.3}
$$

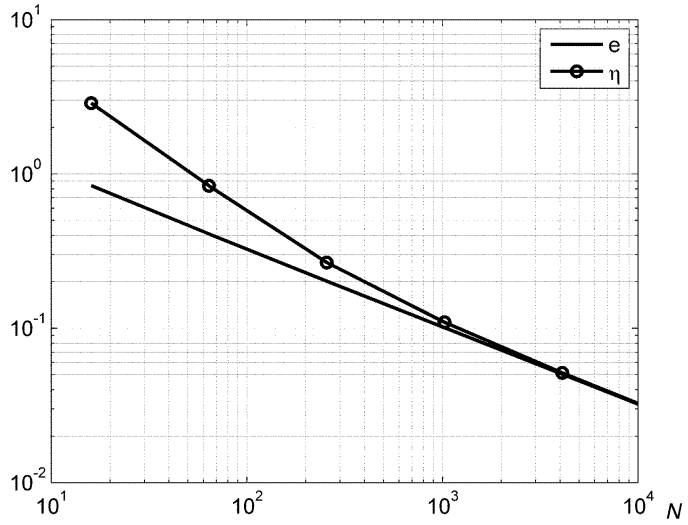

Figure 1: Energy norm of FE error and a posteriori estimator against number of unknowns for the problem (5.1).

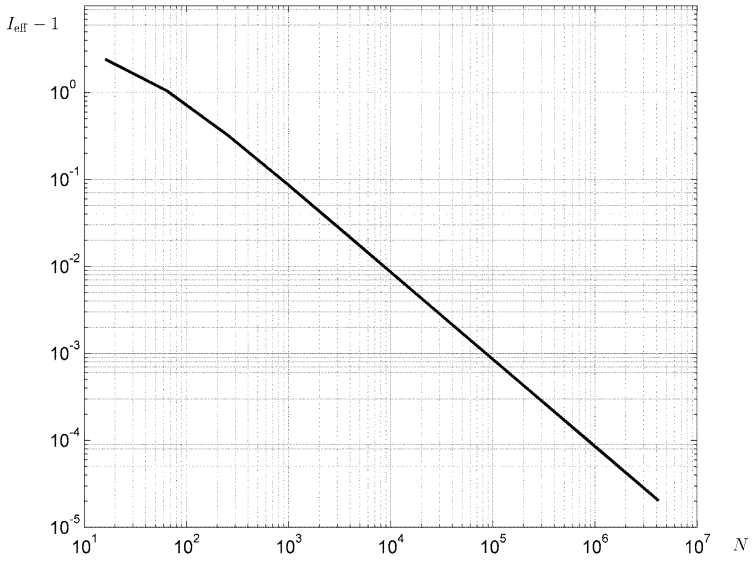

Figure 2: Dependence $I_{\text{eff}} - 1$ on the number of unknowns N for the problem (5.1).

is the effectiveness index of the a posteriori estimate. We see that I_{eff} converges to 1 rather fast and always stays grater then 1, see also Table 1.

We also compared the computational costs of the a posteriori estimator and optimal multigrid solver for the problem (5.1). These results are shown on the Fig. 3 and demonstrate that the equilibrium based a posteriori estimator is optimal with respect to the number of arithmetic operations. Moreover, computation of the a posteriori estimator is about twice cheaper than

solving the finite element system (this results were obtained on AMD Athlon 64 3200+ 2.01GHz with 2 Gb of RAM).

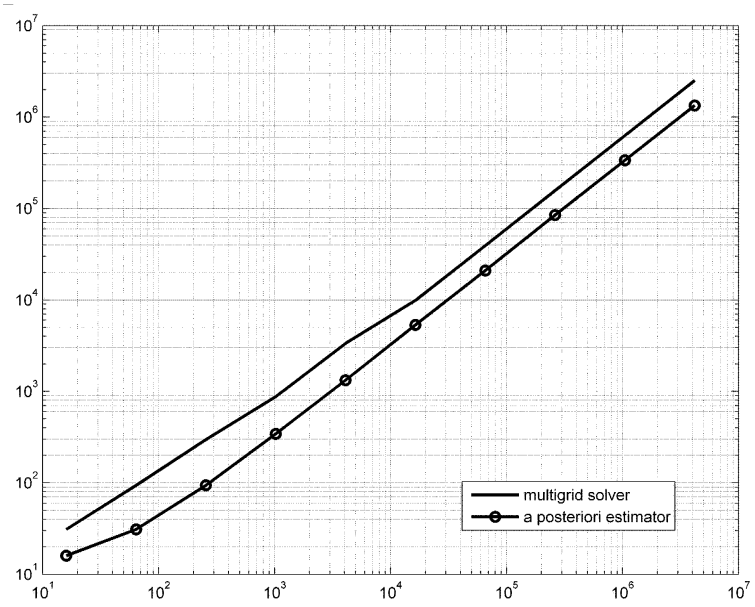

Figure 3: Computational costs (in *ms*) of multigrid solver and η against the number of unknowns for the problem (5.1).

5.1.2 Algorithms based on the Castigliano principle

In the unite square, we considered the Dirichlet problem having for the exact solution

$$u = \sin(2\pi x_1)\sin(\pi x_2) + (x_1+1)(2x_2+1)\,.$$

The corresponding right part is $f = 5\pi^2 \sin(2\pi x_1)\sin(\pi x_2)$ and nonhomogeneous Dirichlet boundary condition is $u\big|_{\partial\Omega} = (x_1+1)(2x_2+1)$.

The FE solution $u_{\rm fem}$ was obtained by means of the space of the continuous piece wise bilinear functions. For obtaining the approximate solution $\mathbf{z}^h$ of the discretized dual problem (4.4), we used the subspace Q_0^h with the local flux basis vectors $\boldsymbol{\phi}^{(i)}$, described in Subsection 4.1. Two choices for the vector $\mathbf{t}_f$ and respectively for the set Q_f^h were implemented: one according to second line of (4.8) and another according to the formulas

$$\mathbf{t}_f(x) = (t_{f,1}, t_{f,2})^\top\,, \qquad t_{f,k}(x) = \frac{1}{2}\int_0^{x_k} f(\eta_k, x_{3-k})\; d\eta_k\;.$$

For these two choices of fluxes, satisfying the balance equation, we introduce notations $\mathbf{t}_f = \mathbf{t}_f^{(1)}, \mathbf{t}_f^{(2)}$. These fluxes result in the two a posteriori estimators

$$\eta_{c,j} = \|\nabla u_{\rm fem} - \mathbf{z}^{h,j}\|\,, \qquad \mathbf{z}^{h,j} = \mathbf{z}_0^{h,j} + \mathbf{t}_f^{(j)}, \qquad j = 1,2\,,$$

with the effectiveness indices denoted $I^{\mathrm{j}}_{\mathrm{eff}}$. The direct evaluation algorithm was also applied to the problem. In essence it is Algorithm 2.1, but adapted to the case, when one has the Dirichlet boundary condition at Γ_N as it is defined in (2.2). The adaptation was performed with the use of special algorithm, which optimizes the a posteriori estimator among different boundary fluxes in the appropriate finite dimensional space. The effectiveness index for the the estimator produced by the adapted Algorithm 2.1 is denoted I_{eff}.

The dependence of the energy norm for the error and of the described effectiveness indices on N is presented in Table 2.

N	e	I_{eff}	I^1_{eff}	I^2_{eff}
16	1.59912	2.93605	2.05747	1.20610
64	$7.60498 \cdot 10^{-1}$	1.95927	2.16500	1.23228
256	$3.70781 \cdot 10^{-1}$	1.26742	2.21586	1.24493
1024	$1.84023 \cdot 10^{-1}$	1.07113	2.23084	1.24869
4096	$9.18344 \cdot 10^{-2}$	1.01826	2.23475	1.24967
16384	$4.58949 \cdot 10^{-2}$	1.00460	2.23574	1.24992
65536	$2.29446 \cdot 10^{-2}$	1.00115	2.23599	1.24998
262144	$1.14720 \cdot 10^{-2}$	1.00029	2.23605	1.24999
1048576	$5.73594 \cdot 10^{-3}$	1.00007	2.23606	1.25000

Table 2.

Other problems were also solved by FE method on purpose of testing the Casigliano principle based a posteriori estimators. The numerical results showed similar behavior of the estimators of this class and allow us to come to the following conclusions.

$\boldsymbol{\alpha}$) The equilibrated fluxes obtained by means of the Castigliano principle provide good error estimators with good effectiveness indices, one of which stays below 1.3.

$\boldsymbol{\beta}$) However, both indices $I^{\mathrm{j}}_{\mathrm{eff}}$ do not converge at $h \to 0$, whereas I_{eff} does.

$\boldsymbol{\gamma}$) The computational time for solving the system of algebraic equations (4.5), resulting from the Castigliano principle, and for evaluating the a posteriori estimators is proportional to the number of unknowns N. Therefore, the algorithms are asymptotically optimal in the computational cost. However, the computer time is greater approximately in 1.5 times, than the computational time for the direct evaluation algorithm with the effectiveness index I_{eff}.

Item $\boldsymbol{\beta}$) can be explained by the fact that the FE method for solving the primal and dual problems have the same rate of convergence in the energy norms. Namely,

$$\|\nabla(u - u_{\mathrm{fem}})\| \asymp \|\nabla u - \mathbf{z}\| = \mathcal{O}(h)\,.$$

Apparently, the convergence of the effectiveness index to the unity for our algorithms of the direct evaluation of the balanced fluxes is related to the fact of the superconvergence of the FE solution u_{fem} to the continuous piece wise bilinear interpolation of the exact solution u.

Item $\boldsymbol{\gamma}$) completely approves conclusions made in Subsection 4.2 in the part titled `Comparison of discrete primal and dual formulations`.

The set Q_f does not depend on the choice of the vector $\mathbf{t}_f$, which enter the definition of this set. However, the sets $Q^{h,j}_f := Q^h_0 + \mathbf{t}^{(j)}_f$ depend, and, according to the numerical results, some

allow to approximate true fluxes better. Besides, the error estimator $\eta_{c,2}$ is more symmetric with respect to the axes x_k, than $\eta_{c,1}$. Probably, these factors caused the difference of the effectiveness indexes, reflected in $\boldsymbol{\alpha}$).

5.2 Second order elliptic equation with discontinuous coefficient

We tested also our a posteriori estimator as applied to the problem:

$$-\nabla \cdot (\rho(x)\nabla u) = f(x), \qquad x \in \Omega = (0,1) \times (0,1)\,, \tag{5.4}$$

with the same boundary conditions as in (5.1) and the piece-wise constant coefficient, which has a jump across the common boundary γ for the two parts of Ω:

$$\Omega_1 = \{x|\ x_1 \in (0,\, 0.5), \quad x_2 \in (0,1)\}\ , \tag{5.5}$$

$$\Omega_2 = \{x|\ x_1 \in (0.5,\, 1), \quad x_2 \in (0,1)\}\ . \tag{5.6}$$

For the function ρ, we used

$$\rho(x) = \begin{cases} \rho_1 = 10^{-2}, & x \in \Omega_1\,, \\ \rho_2 = 10^{2}, & x \in \Omega_2\,. \end{cases} \tag{5.7}$$

The right-hand side f as well as the mixed boundary conditions corresponded to the exact solution

$$u = (\cos(2\pi x) - 1)\cos\left(\frac{3\pi}{2}y\right) \begin{cases} x^2 + 1, & x < 0.5 \\ -(x-1)^2\frac{\rho_1}{\rho_2} + 0.25\frac{\rho_1}{\rho_2} + 1.25, & x > 0.5 \end{cases} \tag{5.8}$$

For obtaining FE solution, we used the space $\mathcal{V}_0(\Omega)$ of the continuous piece-wise bilinear functions, satisfying the Dirichlet boundary condition on Γ_D.

N	e	η	I_{eff}
64	9.14308	$1.02954 \cdot 10^{1}$	1.12604
256	4.44483	4.77467	1.07421
1024	2.20395	2.25318	1.02234
4096	1.09958	1.10628	1.00609
16384	$5.49486 \cdot 10^{-1}$	$5.50364 \cdot 10^{-1}$	1.00160
65536	$2.74705 \cdot 10^{-1}$	$2.74818 \cdot 10^{-1}$	1.00041
262144	$1.37348 \cdot 10^{-1}$	$1.37362 \cdot 10^{-1}$	1.00010
1048576	$6.86734 \cdot 10^{-2}$	$6.86751 \cdot 10^{-2}$	1.00003

Table 3.

The vector $\mathbf{t}(x) = (t(x)_1, t(x)_2)^\top$, satisfying the balance equation (2.5) and the Neumann boundary conditions (5.1) on Γ_N, was calculated according to Algorithm 2.4 in Subsection 2.3. In turn, this vector allowed to evaluate the a posteriori estimator $\eta_\rho = \|\rho\nabla u_{\text{fem}} - \mathbf{t}\|_{\rho^{-1}}$, which enters the estimate (2.33). Figure 4 shows the dependence of the energy norm, of the error

$e = \|\nabla(u - u_{\rm fem})\|_\rho$ and of η_ρ on the number of unknowns N (in this numerical experiment N also exceeded $1 \cdot 10^6$, but for $N > 10^4$ the lines on the graph coincide). The effectiveness index tends to 1 rather fast, as it is illustrated by Fig. 5, in which the value of $I_{\rm eff} - 1$ is plotted against N. At the same time the effectiveness index is always grater then 1 (see also Table 2) that validates that the a posteriori estimate (2.33) is a *guaranteed upper asymptotically exact bound*.

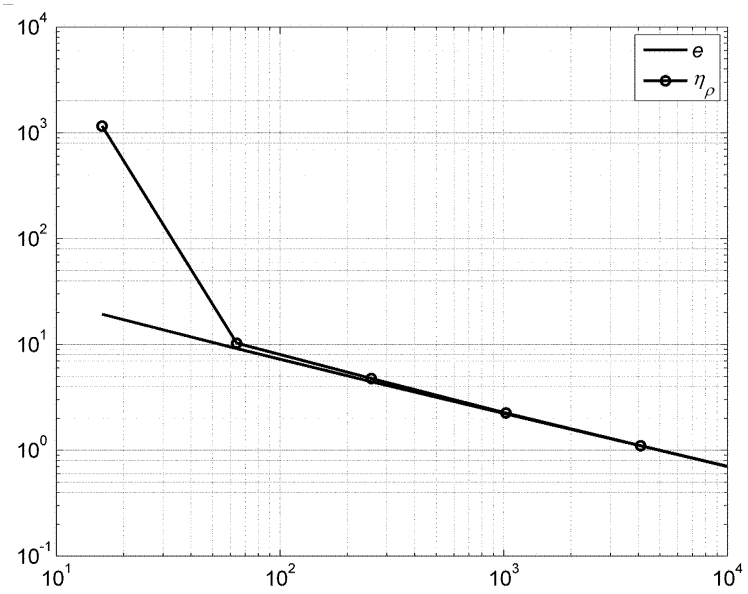

Figure 4: The energy norm of the finite element error and the a posteriori estimator against the number of unknowns for the problem (5.4).

The comparison of the computational costs of the a posteriori estimator and the optimal multigrid solver for the FE system of linear algebraic equations is presented in Fig. 6. These results demonstrate the optimality of the a posteriori estimator. Note, that the a posteriori estimator is more than twice cheaper, than solving the FE system by the multigrid method.

5.3 Linear elasticity problem

Algorithms 3.1 and 3.2 were applied to several linear elasticity problems with nonhomogeneous Dirichlet boundary conditions. Since the results reflect similarly the difference of these algorithms and the level of their efficiency, we present them only for one problem (3.1)-(3.3) in the unite square $\Omega = [0,1]\times[0,1]$. The vector $\mathbf{f}$ and the Dirichlet boundary conditions correspond to the exact solution $\mathbf{u} = (u_1, u_2)^\top$,

$$\begin{aligned} u_1(x) &= \sin(\pi x_1)\sin(2\pi x_2) + x_1 + x_2 \ , \\ u_2(x) &= \sin(2\pi x_1)\sin(\pi x_2) + \tfrac{1}{4}(x_1+1)(x_2+1) \ . \end{aligned} \tag{5.9}$$

Algorithms 3.1 and 3.2 produce the a posteriori estimators $\eta^{(1)}$ and $\eta^{(2)}$, respectively, according to their description in Subsection 3.3.

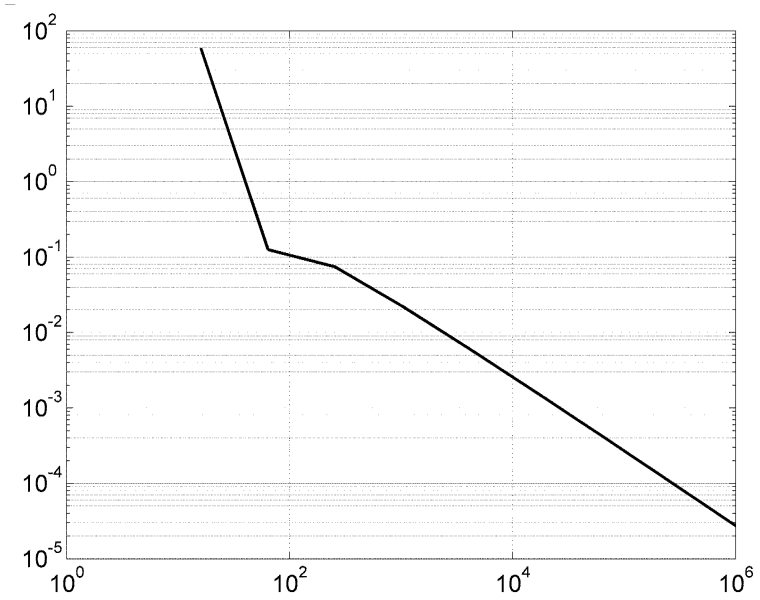

Figure 5: Dependence of $I_{\text{eff}} - 1$ on the number of unknowns N for the problem (5.4).

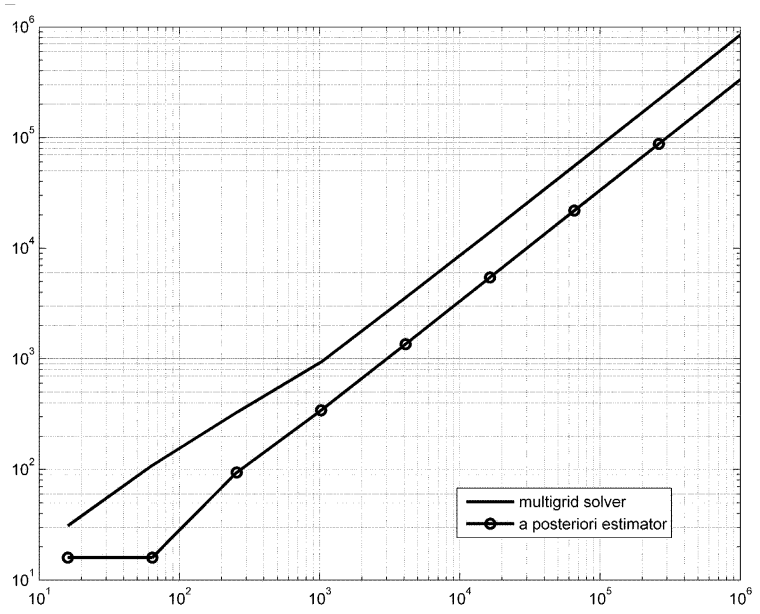

Figure 6: The computational costs of the a posteriori estimator and the multigrid solver against the number of unknowns for the problem (5.4).

Figures 7 and 8 demonstrate the behavior of $\eta^{(1)}$. The results obtained for $\eta^{(2)}$ are presented in Figures 9-11, see also Table 3 for numbers. The numerical results show, that the a posteriori estimator $\eta^{(2)}$ outperforms $\eta^{(1)}$. The effectiveness index $I_{\text{eff}}^{(2)}$ of $\eta^{(2)}$ tends to 1 staying grater then 1, whereas the effectiveness index $I_{\text{eff}}^{(1)}$ of $\eta^{(1)}$ does not converge and stays slightly greater than 2. Both a posteriori estimators are optimal in the computational cost. Figures 8 and 11 present

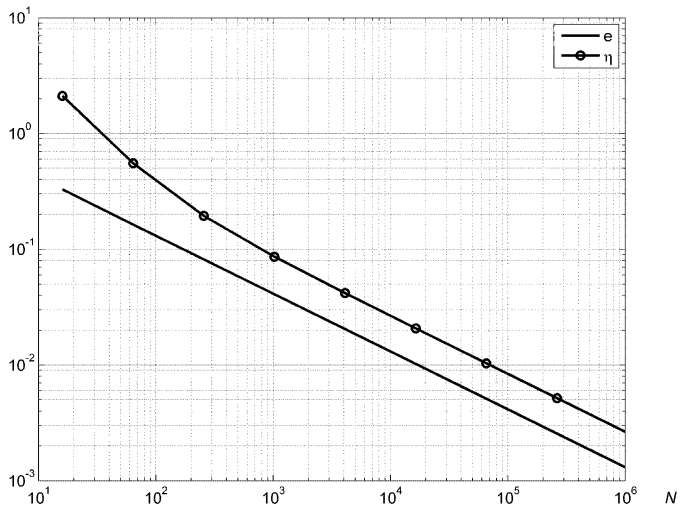

Figure 7: The energy norm of FE error and a posteriori estimator $\eta^{(1)}$ against the number of unknowns for the problem (5.9)

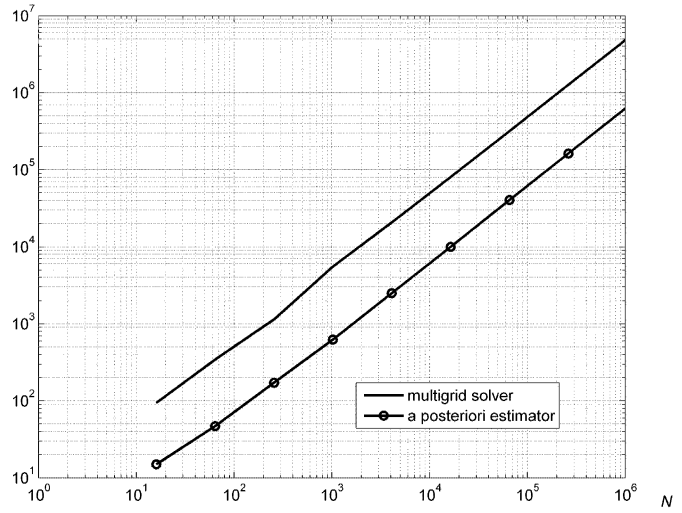

Figure 8: Computational cost of multigrid solver and of the a posteriori estimator $\eta^{(1)}$ against the number of unknowns for the problem (5.9).

the comparison of their computational cost and the computational cost of the multigrid solver for solving the problem (3.1)-(3.3) with the Dirichlet boundary conditions.

The difference in the behavior of the a posteriori estimators $\eta^{(1)}$ and $\eta^{(2)}$ shows that the quality of the error estimators strongly depend on how accurately the properties of the superconvergence of FE solutions are taken into account at their evaluation. Clearly, *the convergence of the effectiveness index can take place only if an a posteriori error estimator is superconver-*

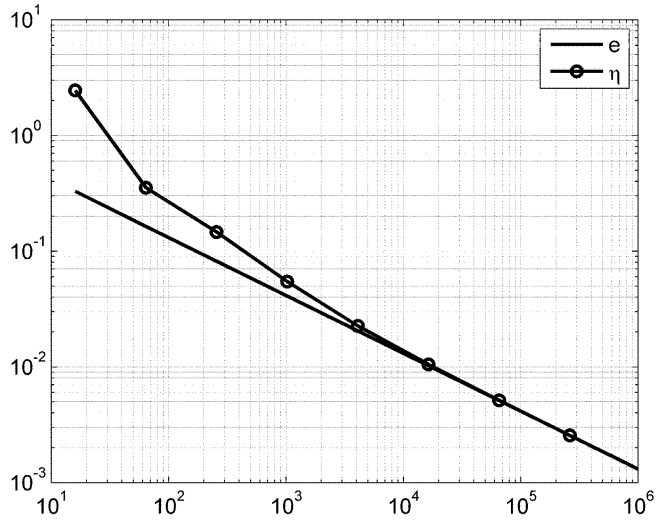

Figure 9: Energy norm of FE error and the a posteriori estimator $\eta^{(2)}$ against the number of unknowns for the problem (5.9).

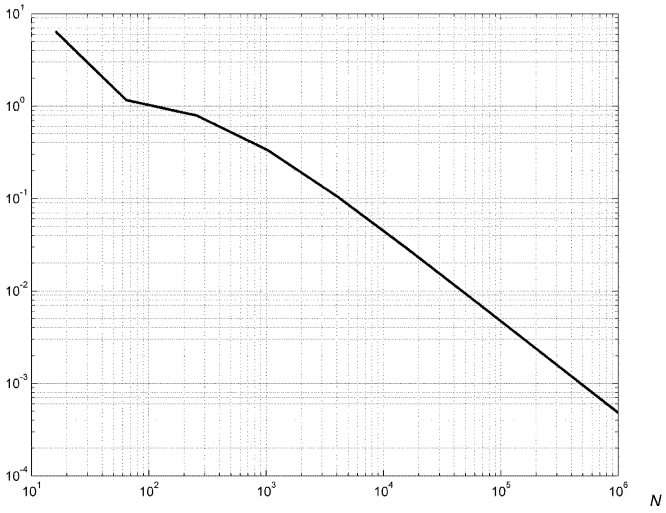

Figure 10: Dependence of $I_{\text{eff}} - 1$ for the aposteriori estimator $\eta^{(2)}$ on the number of unknowns N for the problem (5.9).

gent. Algorithm 3.1 is simpler, but it is not invariant in respect to axes x_k and the accuracy of approximation of the boundary values for stresses provides only an approximation of the order of h.

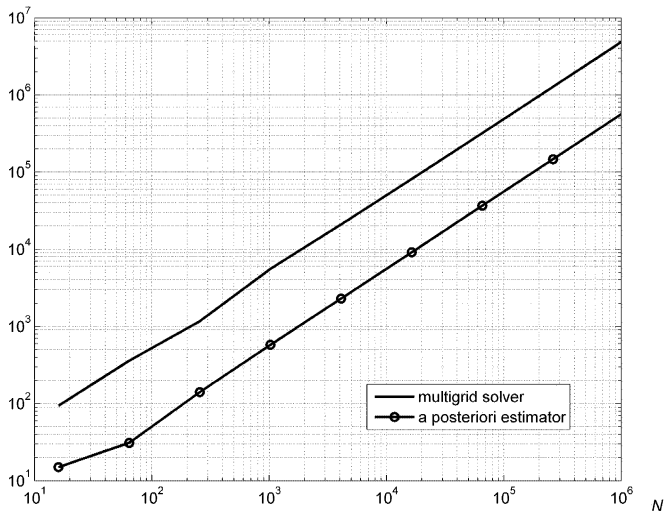

Figure 11: Computational costs of the multigrid solver and the a posteriori estimator $\eta^{(2)}$ against the number of unknowns for the problem (5.9).

N	e	$\eta^{(1)}$	$I_{\text{eff}}^{(1)}$	$\eta^{(2)}$	$I_{\text{eff}}^{(2)}$
16	$3.29126 \cdot 10^{-1}$	2.11430	6.42399	2.45022	7.44462
64	$1.63481 \cdot 10^{-1}$	$5.54014 \cdot 10^{-1}$	3.38885	$3.54162 \cdot 10^{-1}$	2.16638
256	$8.17444 \cdot 10^{-2}$	$1.94481 \cdot 10^{-1}$	2.37914	$1.46349 \cdot 10^{-1}$	1.79033
1024	$4.08766 \cdot 10^{-2}$	$8.62833 \cdot 10^{-2}$	2.11083	$5.45946 \cdot 10^{-2}$	1.33560
4096	$2.04389 \cdot 10^{-2}$	$4.18226 \cdot 10^{-2}$	2.04622	$2.56000 \cdot 10^{-2}$	1.10387
16384	$1.02196 \cdot 10^{-2}$	$2.07433 \cdot 10^{-2}$	2.02976	$1.05054 \cdot 10^{-2}$	1.02797
65536	$5.10979 \cdot 10^{-3}$	$1.03501 \cdot 10^{-2}$	2.02554	$5.14644 \cdot 10^{-3}$	1.00717
262144	$2.55490 \cdot 10^{-3}$	$5.17229 \cdot 10^{-3}$	2.02446	$2.55953 \cdot 10^{-3}$	1, 00181
1048576	$1.27745 \cdot 10^{-3}$	$2.58580 \cdot 10^{-3}$	2.02419	$1.27804 \cdot 10^{-3}$	1.00046

Table 4.

References

[1] Ainsworth M, Demkowicz L and Kim C-W. Analysis of the equilibrated residual method for a posterioti estimation on meshes with hanging nodes. Submitted for publication.

[2] Ainsworth M and Oden JT. *A posteriori estimation in finite element analysis*. John Wiley & Sons, Inc., New York, 2000: xx+240p.

[3] Aubin J-P. *Approximation of elliptic boundary-value problems.* Wiley Interscience, a division of John Wiley & Sons, Inc., New York-London-Sydney-Toronto, 1972.

[4] Arthurs AM. *Complementary variational principles.* Calderon Press, Oxford, 1980.

[5] Babuška I and Reinbolt WC. Error estimates for adaptive finite element computations. *SIAM J. Num. Anal.*, 1978, **15** (4): 736-754.

[6] Babuška I and Reinbolt WC. A posteriori error analysis for finite element methods for one dimensional problems. *SIAM J. Num. Anal.*, 1981, **13** (3): 565-589.

[7] Babuška I and Strouboulis T. *Finite element method and its reliability.* Oxford University Press, New York, 2001: xi+802p.

[8] Berdichevskii VL. *Variatsionnye principy mahaniki sploshnoi sredy (Variational principles of mechanics of conytinuous media).* Nauka, Moskow, 1983. (In Russian)

[9] Bernardi Ch. and Girault V. A local regularization operator for triangular and quadrilateral finite elements. *SIAM J. Num. Anal.*, 1998, **35** (5): 1893-1916.

[10] Braess D and Schöberl J. Equilibrated residual error estimator for Maxswell's equations. Submitted for publication.

[11] Carstensen C. Residual-based a posteriori error estimate for a nonconforming Reissner-Mindlin plate finite element. *SIAM J. Num. Anal.*, 2002, **39** (6): 2034-2044.

[12] Clèment Ph. Approximation by finite element functions using local regularization. *Rev. Francaise Automat. Informat. Recherche Opèrationelle Sèr. Rouge Anal. Numèr.,* 1975, **9** (2): 72-84.

[13] Duvaut G and Lions I-I. *Inequalities in mechanics and physics.* Springer-Verlag, Berlin, Hedelberg, New York, 1976.

[14] Ekeland I and Temam R. *Convex analysis of variational problems.* North-Holland, Amsterdam, New York, Oxford, 1976.

[15] Fraeijs de Vebeke. Displacement and equilibrium models in the finite element method. In: *Stress Analysis*, OC Zienkiewicz and GS Holister eds, Wiley, London-New York, 1965: 145-197.

[16] Glowinski R. *Numerical methods for nonlinear variational problems.* Springer-Verlag, New York-Berlin-Heidelberg-Tokio, 1984.

[17] Gol'denweizer AL. *Theory of elastic thin shells (Translation of original Russian edition of 1953).* Pergamon Press, New York, 1961.

[18] Kelly D. The self equilibration of residuals and complementary a posteriori error estimates in the finite element method. *Internat. J. Num. Methods Engrg.,* 1984 **20**: 1491-1506.

[19] Korneev VG. Primenenie metoda mehanicheskih kvadratur dlia postroenia raznostnoi shemy dlia ellipticheskogo uravnenia chetvertogo poriadka (Implementation of quadrature method for costruction of difference method for elliptic equation of 4-th order). *Metody vychislenii*, 1971, N7, Isdatel'stvo Leningradskogo universiteta, Leningrad: 46-55. (In Russian)

[20] Korneev VG. O sviazi raznostnyh shem dlia ellipticheskih uravnenii vtorogo poriadka so shemami kvadratur dlia integral'nyh uravnenii metoda raschlenenia (On interrelation between difference schemes for elliptic equations of second order and quadrature method for integral equations of splitting method). *Vestnik Leningradskogo universiteta*. 1973, (1): 18-27. (In Russian)

[21] Korneev VG. Raznostnaya shema poriadka tochnosti $h^{3/2}$ v norme $\|\cdot\|_{h,W_2^1}$ dlia ellipticheskogo uravnenia vtorogo poriadka v proizvol'noi oblasti (Difference scheme of accuracy $h^{3/2}$ in the norm $\|\cdot\|_{h,W_2^1}$ for elliptic equation of second order in arbitrary domain). *Vestnik Leningradskogo universiteta*. 1973, (7): 25-33. (In Russian)

[22] Korneev VG. Svedenie zadachi rascheta arochnoi plotiny k zadache na minimum raboty uprugih sil nepreryvnoi sterzhnevoi sistemy (Reduction of the arch dam analysis problem to minimization of work of elastic forces for continuous rod system). *Trudy Gidroproekta*, pod redaktsiei D.M. Yurinova. 1973, N 22, Moskva: 4-12. (In Russian)

[23] Korneev VG. Diskretizatsia raboty uprugih sil nepreryvnoi sterzhnevoi sistemy (Discretization of functional of elastic forces work for continuous rod system). *Trudy Gidroproekta*, pod redaktsiei D.M. Yurinova. 1973, N 22, Moskva: 12-32. (In Russian)

[24] Korneev VG. O ratsional'nom vybore koordinatnyh funktsii pri minimizatsii diskretnogo funktsionala raboty uprugih sil (On rational choice of the coordinate functions at minimization of the discrete functional of the work of elastic forces). *Trudy Gidroproekta*, pod redaktsiei D.M. Yurinova. 1973, N 22, Moskva: 86-102. (In Russian)

[25] Korneev VG. Issledovanie shodimosti priblizhennyh reshenii (Analysis of convergence of approximate solutions). *Trudy Gidroproekta*, pod redaktsiei D.M. Yurinova, 1973, N 22, Moskva: 141-182. (In Russian)

[26] Korneev VG. *Shemy metoda konechnyh elementov poriadkov tochnosti (The finite elelment methods of high oreders of accuracy)*. Izdatel'stvo Leningradskogo universiteta, Leningrad, 1977. (In Russian)

[27] Korneev VG. On numerical solution of the shell theory problems in stresses on skew-angular meshes. *Comput. Maths and Math. Physics,* 1981, **21** (2): 441-450.

[28] Korneev VG. Superconvergence of the finite element solutions in the mesh norms. *Comput. Maths and Math. Physics,* 1982, **22** (5): 1133-1148.

[29] Korneev VG and Rozin LA. Algorithmy rascheta obolochek po sterzhnevoi sheme i ih matematicheskoe issledovanie (Algorithms for numerical solution of shell problems and their mathematical analysis). In: *Materialy 6-oi vsesoyuznoi konferetsii po teorii obolochek i plastin (Proc. of 6-th all-union conference on shell and plate theory)* , Baku, 1966, Nauka: 555-560. (In Russian)

[30] Korneev VG and Rozin LA. Chislennaia relizatsia metoda raschlenenia na primere zadachi izgiba plastinki (Numerical realization of the splitting method for the plate bending problem). *Izvestia akademii nauk SSSR, Mehanika tverdogo tela.* 1967 (2). (In Russian)

[31] Ladevese P and Leguillon D. Error estimate procedure in the finite element method and applications. *SIAM J. Num. Anal.*, 1983, **20**: 485-509.

[32] Luce R and Wohlmuth B. A local a posteriori error estimator based on equilibrated fluxes. *SIAM J. Num. Anal.*, 2004, **42**: 1394-1414.

[33] Mikhlin SG. *Variational methods in mathematical physics.* Pergamon, Oxford, 1964.

[34] Mosolov PP and Myasnikov VP. *Mechanics of rigid plastic bodies.* Nauka, Moskow, 1981. (In Russian)

[35] Neittaanmäki P and Repin SI. *Reliable methods for computer simulation Error control and a posteriori estimates.* Elsevier, New York, 2004, 305p.

[36] Novozhilov VV. *Teoria tonkih obolochek (Theory of thin shells).* Gosudarstvennoe soyuznoe izdatel'stvo sudostroitel'noi promyshlennosti, Leningrad, 1962, 431p. (In Russian)

[37] Novozhilov VV. *Theory of thin shells.* Noordoff, Dordrecht, 1959.

[38] Oganesian LA and Ruhovets LA. *Variatsionno-raznostnye metody reshenia ellipticheskih uravnenii (Variational-difference methods for solution of elliptic equations).* Izdatel'stvo Armianskoi akademii nauk, Erevan, 1979, 234p.

[39] Oden JT, Demkowicz L, Rachowicz W and Westermann TA. Towards a universal h-p adaptive finite element strategy. Part 2, A posteriori error estimation. *Comput. Methods Appl. Mech. Engrg.,* 1989, **77**: 113-180.

[40] Oden T, Duarte C. and Zienkiewcz O. A new cloud based finite element method. *Comput. Methods Appl. Mech. Engrg.*, 1998, **153**: 117-126.

[41] Repin S and Frolov M. Ob a posteriornyh otsenkah tochnosti priblizhennyh reshenii kraevyx zadach dlia ellipticheskih uravnenii. (On a posteriori estimates of approximate solutions of boundary value problems for elliptic equations). *Zhurnal vychislit. matem. i matem. phiziki*, 2002, **42** (12): 1774-1787.

[42] Repin S, Neittaanmäkki P and Frolov M. On computational properties of a posteriori esimates based upom the method of duality error majorants. In: *Numerical Mathematics and Advanced Applications (ENUMATH-2003).* Berlin-Heidelberg, Springer-Verlag, 2004: 346-357.

[43] Repin SI and Xanthis LS. A posteriori error estimation for elasto-plasric problems based on duality theory. *Comput. Meth. Appl. Mech. Engrg.,* 1996, **138**: 317-339.

[44] Reissner E. On the derivation of the theory of thin elastic shells. *J. Math. Phis*, 1963, **42** (4).

[45] Rozin LA. Metod raschlenenia v teorii obolochek (The splitting method in the shell theory). *Prikladnaia matematika i mehanika*, 1961, **25** (5). (in Russian)

[46] Rozin LA. Nekotorye voprosy raschlenenia i diskretizatsii uravnenii teorii obolochek (Some questions of splitting and discretization of equations of the shell theory). *Issledovania po teorii uprugosti i plastichnosti.* 1965 (4), Izdatel'stvo Leningradskogo universiteta, Leningrad.

[47] Strouboulis T, Babuska I and Copps K. The design and analysis of the generalized finite element method. *Comput. Methods Appl. Mech. Engrg.*, 2000, **181**: 43-69.

[48] Stewart JR Hudges TJR. A tutorial in elementary finite element error analysis. A systematic presentation of a priori and a posteriori error estimates. *Comput. Meth. Appll. Mech. Eng.,* 1998, **158** (1-2): 1-22.

[49] Temam R. *Problèmes Mathématique en Plasticité*. Bordas, Paris, 1983.

[50] Vejchodský T. Local a posteriori error estimator based on the hypercircle method. In: *Proc. of the European Congress on Computational Method in Applied Mechanics and Engrg. (ECCOMUAS 2004).* Yavaskylä, Finland.

[51] Verfürth RA. A posteriori error estimates for nonlinear problems. Finite element discretizations of elliptic equations. *Math. of Comput.*, 1994, **62** (206): 445-465.

[52] Verfürth RA. *A review of a posteriori error estimation and adaptive mesh refinement techniques.* John Wiley & Sons, BG Teubner, 1996: vi+127.

[53] Walbin LB. *Superconvergence in Galerkin finite elelemnt methods.* Lecture notes in Mathematics. 1995, Springer, Berlin.

[54] Washizu K. *Variational principles in elasticity and plasticity, 3-rd edition.* Pergamon Press, New York, 1982

[55] Zienkiewicz OC and Zhu JZ. The superconvergence path recovery (SPR) and adaptive finite element refinement. *Comput. Meth. Appll. Mech. Eng.*, 1992 **29** (1): 78-88.